

BIOS220- INTRODUCTION TO GENETICS



Jennifer Yeack
University of Illinois Chicago

BIOS220- Introduction to Genetics

Jennifer Yeack

University of Illinois Chicago

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (<https://LibreTexts.org>) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by [NICE CXone Expert](#) and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptations contact info@LibreTexts.org or visit our main website at <https://LibreTexts.org>.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 08/23/2026

TABLE OF CONTENTS

Licensing

1: Module 1

1.1: What is DNA?

- 1.1.1: Historical Basis of Modern Understanding
- 1.1.2: The Structure of DNA
- 1.1.3: Genome Organization

1.2: DNA Replication

- 1.2.1: Basics of DNA Replication
- 1.2.2: DNA Replication
- 1.2.3: DNA Replication in Bacteria
- 1.2.4: DNA Replication in Eukaryotes

1.3: Transcription

- 1.3.1: Transcription
- 1.3.2: Bacterial Transcription
- 1.3.3: Eukaryotic Transcription
- 1.3.4: RNA Processing

2: Module 2

2.1: Mitosis, Meiosis, and the Cell Cycle

- 2.1.1: The Cell Cycle
- 2.1.2: Control of the Cell Cycle
- 2.1.3: Cancer and the Cell Cycle
- 2.1.4: Prokaryotic Cell Division
- 2.1.5: Meiosis
- 2.1.6: Errors in Meiosis

2.2: Translation

- 2.2.1: The Genetic Code
- 2.2.2: Ribosomes and Protein Synthesis

2.3: Gene Expression

- 2.3.1: Regulation of Gene Expression
- 2.3.2: Prokaryotic Gene Regulation
- 2.3.3: Eukaryotic Gene Regulation
- 2.3.4: Cancer and Gene Regulation

2.4: Mutations and DNA Repair

- 2.4.1: DNA Damage and Repair
- 2.4.2: Types of Mutations

3: Module 3

3.1: Inheritance

- 3.1.1: Mendel's Experiments
- 3.1.2: Laws of Inheritance
- 3.1.3: Extensions of the Laws of Inheritance
- 3.1.4: Black fur color - a dominant trait
- 3.1.5: Yellow fur color - a recessive trait
- 3.1.6: Brindle color - partial dominance and epistasis
- 3.1.7: White spotting - when there's more than two alleles
- 3.1.8: KIT - embryonic lethality
- 3.1.9: Hemophilia - a sex-linked disorder
- 3.1.10: It's not all in the genes - the effect of environment
- 3.1.11: Pleiotropy - one gene affects more than one trait

4: Genetic Technology

- 4.1: Biotechnology
- 4.2: Mapping Genomes
- 4.3: Whole-Genome Sequencing
- 4.4: Genomics and Proteomics**
 - 4.4.1: Biotechnology

Index

Glossary

Detailed Licensing

Licensing

A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.

1: MODULE 1

1.1: What is DNA?

1.1.1: Historical Basis of Modern Understanding

1.1.2: The Structure of DNA

1.1.3: Genome Organization

1.2: DNA Replication

1.2.1: Basics of DNA Replication

1.2.2: DNA Replication

1.2.3: DNA Replication in Bacteria

1.2.4: DNA Replication in Eukaryotes

1.3: Transcription

1.3.1: Transcription

1.3.2: Bacterial Transcription

1.3.3: Eukaryotic Transcription

1.3.4: RNA Processing

1.1: WHAT IS DNA?

1.1.1: Historical Basis of Modern Understanding

1.1.2: The Structure of DNA

1.1.3: Genome Organization

This page titled [1.1: What is DNA?](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by Jennifer Yeack.

1.1.1: Historical Basis of Modern Understanding

Today we know DNA, or deoxyribose nucleic acid, functions as the genetic material. Our current understanding of DNA began with the discovery of nucleic acids followed by the development of the double-helix model. In the 1860s, Friedrich Miescher, a physician by profession, isolated phosphate-rich chemicals from white blood cells (leukocytes). He named these chemicals (which would eventually be known as DNA) *nuclein* because they were isolated from the nuclei of the cells.

A half century later, in 1928, British bacteriologist **Frederick Griffith** reported the first demonstration of **bacterial transformation**—a process in which external DNA is taken up by a cell, thereby changing its morphology and physiology. Griffith conducted his experiments with *Streptococcus pneumoniae*, a bacterium that causes pneumonia. Griffith worked with two strains of this bacterium called rough (R) and smooth (S) (the two cell types were called “rough” and “smooth” after the appearance of their colonies grown on a nutrient agar plate).

The R strain is non-pathogenic (does not cause disease). The S strain is pathogenic (disease-causing), and has a capsule outside its cell wall. The capsule allows the cell to escape the immune responses of the host.

When Griffith injected the living S strain into mice, they died from pneumonia. In contrast, when Griffith injected the live R strain into mice, they survived. In another experiment, he heat-killed a culture of the S strain. When he injected mice with this heat-killed S strain, they also survived. In a third set of experiments, a mixture of live R strain and heat-killed S strain were injected into mice, and—to his surprise—the mice died. Upon isolating the live bacteria from the dead mouse, only the S strain of bacteria was recovered. When this isolated S strain was injected into fresh mice, the mice died. Griffith concluded that something had passed from the heat-killed S strain into the live R strain and transformed it into the pathogenic S strain. He called this the *transforming principle* (Figure 1.1). These experiments are now known as **Griffith's transformation experiments**.

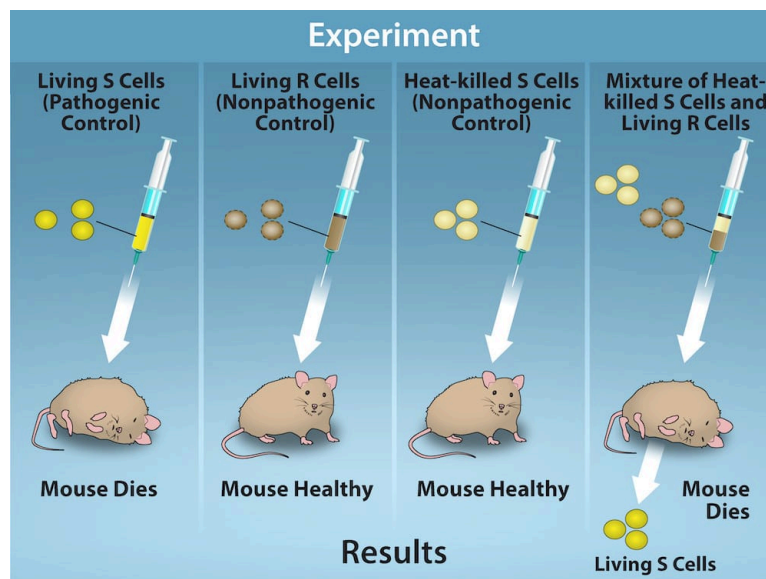


Figure 1.1 Two strains of *S. pneumoniae* were used in Griffith's transformation experiments. The R strain is non-pathogenic, whereas the S strain is pathogenic and causes death. When Griffith injected a mouse with the heat-killed S strain and a live R strain, the mouse died. The S strain was recovered from the dead mouse. Griffith concluded that something had passed from the heat-killed S strain to the R strain, transforming the R strain into the S strain in the process. Credit: Rao, A., Ryan, K. and Tag, A. Department of Biology, Texas A&M University.

Scientists **Oswald Avery, Colin MacLeod, and Maclyn McCarty** (1944) were interested in exploring this transforming principle further. They used the S strain cultured from the dead mice and isolated the proteins and nucleic acids (both RNA and DNA) as these were all possible candidates at the time for the molecule of heredity. To determine which molecule it was, they used a process of elimination by treating the extract with specific digestive enzymes. Protease was used to destroy proteins, RNase was used to destroy RNA, and DNase was used to destroy DNA. They added each enzyme to a separate sample of the S-strain extract, then mixed the treated extract with live R-strain bacteria to see if transformation would still occur. They found that when DNA was

degraded, the resulting mixture was no longer able to transform the bacteria, whereas all of the other combinations were able to transform the bacteria. This led them to conclude that DNA was the transforming principle.

Although the experiments of Avery, McCarty and McLeod had demonstrated that DNA was the informational component transferred during transformation, DNA was still considered to be too simple a molecule to carry biological information. Proteins, with their 20 different amino acids, were still regarded as more likely candidates. The decisive experiment, conducted by **Martha Chase and Alfred Hershey** in 1952, provided evidence that DNA was indeed the genetic material and not proteins. Chase and Hershey were studying a **bacteriophage**—a virus that infects bacteria. Viruses typically have a simple structure: a protein coat, called the capsid, and a nucleic acid core that contains the genetic material (either DNA or RNA) (Figure 1.2). The bacteriophage infects the host bacterial cell by attaching to its surface, and then it injects its nucleic acids inside the cell. The phage DNA makes multiple copies of itself using the host machinery, and eventually the host cell bursts, releasing a large number of bacteriophages.

Hershey and Chase selected radioactive elements that would specifically distinguish the protein from the DNA in infected cells. They labeled one batch of phage with radioactive sulfur, ^{35}S , to label the protein coat. Another batch of phage were labeled with radioactive phosphorus, ^{32}P . Because phosphorus is found in DNA, but not protein, the DNA and not the protein would be tagged with radioactive phosphorus. Likewise, sulfur is absent from DNA, but present in several amino acids such as methionine and cysteine.

Each batch of phage was allowed to infect the cells separately. After infection, the phage bacterial suspension was put in a blender, which caused the phage coat to detach from the host cell. Cells exposed long enough for infection to occur were then examined to see which of the two radioactive molecules had entered the cell. The phage and bacterial suspension was spun down in a centrifuge. The heavier bacterial cells settled down and formed a pellet, whereas the lighter phage particles stayed in the supernatant. In the tube that contained phage labeled with ^{35}S , the supernatant contained the radioactively labeled phage, whereas no radioactivity was detected in the pellet. In the tube that contained the phage labeled with ^{32}P , the radioactivity was detected in the pellet that contained the heavier bacterial cells, and no radioactivity was detected in the supernatant. Hershey and Chase concluded that it was the phage DNA that was injected into the cell and carried information to produce more phage particles, thus providing evidence that DNA was the genetic material and not proteins (Figure 1.3)

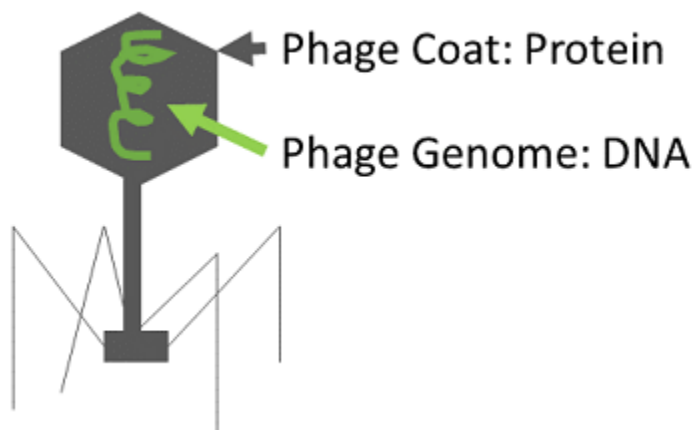


Figure 1.2 Structure of a phage virus. By Angelahartsock - Own work, CC0

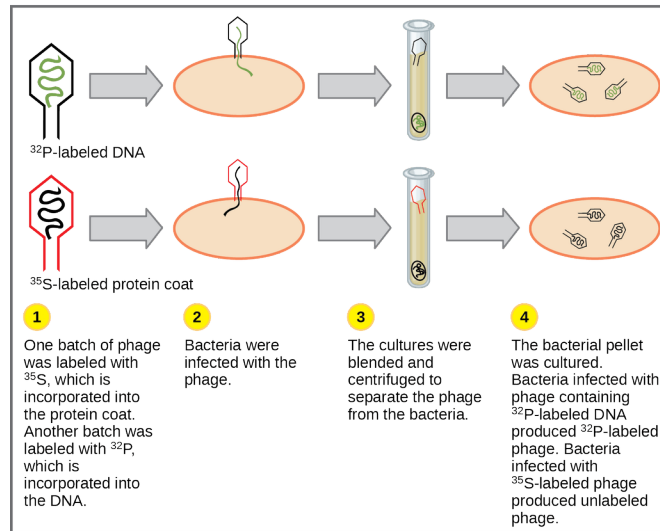


Figure 1.3 In Hershey and Chase's experiments, bacteria were infected with phage radiolabeled with either ^{35}S , which labels protein, or ^{32}P , which labels DNA. Only ^{32}P entered the bacterial cells, indicating that DNA is the genetic material.

Note

Think of radioactive labeling as attaching a microscopic, glowing "GPS tracker" to a specific molecule so scientist can track exactly where it goes. Most elements in nature (like carbon, hydrogen, and phosphorus) exist in a stable form. However, scientists can synthesize versions of these elements with unstable nuclei, known as **radioactive isotopes**. These unstable isotopes naturally emit tiny, harmless bursts of energy (radiation) that can be easily detected using specialized laboratory instruments. By feeding a cell or virus a specific radioactive isotope, scientists can force it to build its molecules using those "glowing" atoms. In the Hershey-Chase experiment, this technique allowed the researchers to look at a mixture of millions of viruses and see exactly where the DNA went versus where the protein went, even though the molecules themselves were invisible under a standard microscope.

Around this same time, Austrian biochemist **Erwin Chargaff** examined the content of DNA in different species and found that the amounts of adenine, thymine, guanine, and cytosine were not found in equal quantities, and that relative concentrations of the four nucleotide bases varied from species to species, but not within tissues of the same individual or between individuals of the same species. He also discovered something unexpected. The amount of adenine equaled the amount of thymine, and the amount of cytosine equaled the amount of guanine (that is, **A = T** and **G = C**). Different species had equal amounts of *purines* (A+G) and *pyrimidines* (T + C), but different ratios of A+T to G+C. These observations became known as Chargaff's rules. Chargaff's findings proved immensely useful when Watson and Crick were getting ready to propose their DNA double helix model! You can see after reading this page of discoveries how science builds upon previous discoveries, sometimes in a slow and laborious process.

Key Terms

Transformation

The uptake and incorporation of foreign DNA by a cell.

Virulent

Capable of causing disease.

Bacteriophage

A virus that infects bacteria.

Transforming principle

The molecule responsible for transferring hereditary traits between cells.

Radioactive isotope

An unstable form of a chemical element that releases energy (radiation) as its nucleus breaks down over time.

Check Your Understanding

1. What was the difference between the R and S strains used by Griffith?
2. Why did mice die when injected with living R bacteria mixed with heat-killed S bacteria?
3. Why did Avery, MacLeod, and McCarty use DNase, RNase, and protease?
4. Why were radioactive phosphorus and sulfur used in the Hershey-Chase experiment?
5. What evidence showed that DNA, not protein, entered bacterial cells?

1.1.2: The Structure of DNA

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, without her knowledge (Figure 1.4). Watson and Crick also had key pieces of information available from other researchers such as Erwin Chargaff. 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. Unfortunately, by then Franklin had died, and Nobel prizes are not awarded posthumously.



(a)



(b)

Figure 1.4 Pioneering scientists (a) James Watson and Francis Crick are pictured here with American geneticist Maclyn McCarty. Scientist Rosalind Franklin discovered 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 nucleic acids are nucleotides, which are made up of three parts:

1. A pentose (5-carbon) sugar. RNA uses ribose and DNA uses deoxyribose
2. A phosphate group. The phosphate group is negatively charged.
3. A nitrogenous base (Figure 1.5).

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**. RNA uses the pyrimidine base uracil (U) instead of thymine. The nucleotide is named according to the nitrogenous base it contains.

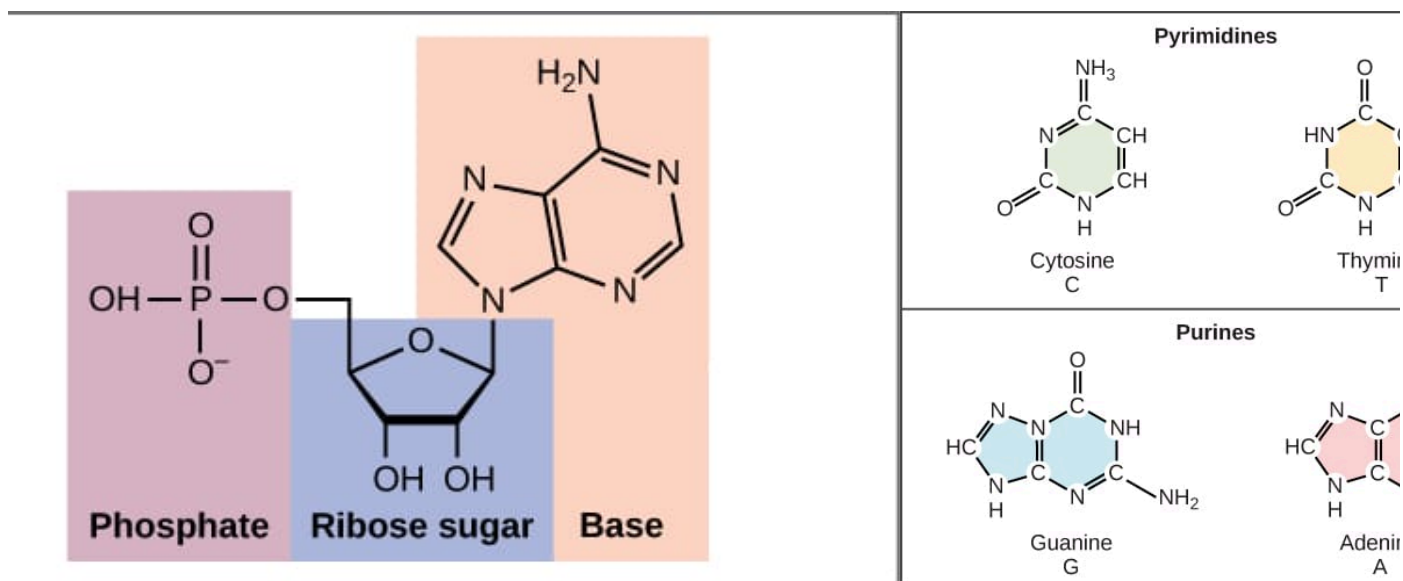


Figure 1.5 (a) Each DNA nucleotide is made up of a sugar, a phosphate group, and a base. (b) Cytosine and thymine are pyrimidines. Guanine and adenine are purines.

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 nitrogenous base is attached to the 1' C. The phosphate group is attached to the 5' carbon, and a hydroxyl (-OH) group is attached to the 3' carbon. The phosphate group of one nucleotide bonds covalently with the 3' carbon of 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. Each DNA molecule is composed of two single strands held together by hydrogen bonds between the bases.

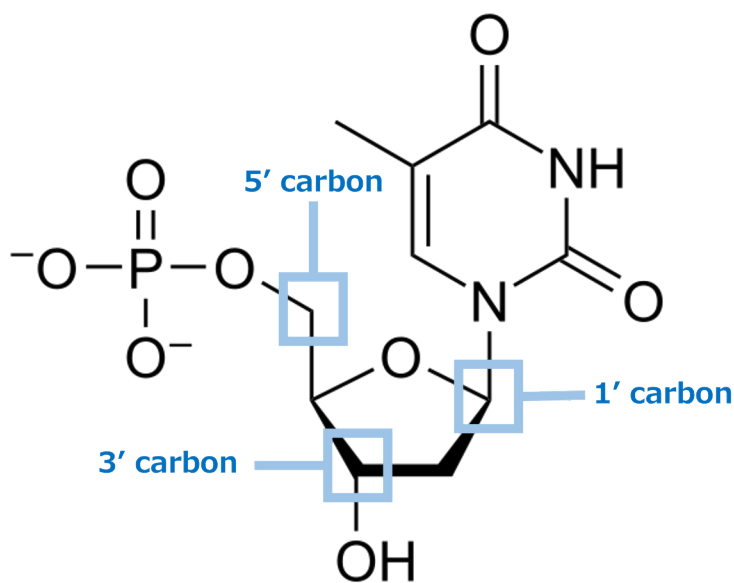


Figure 1.6 The carbon numbering on a nucleotide pentose sugar.

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 (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 1.7).

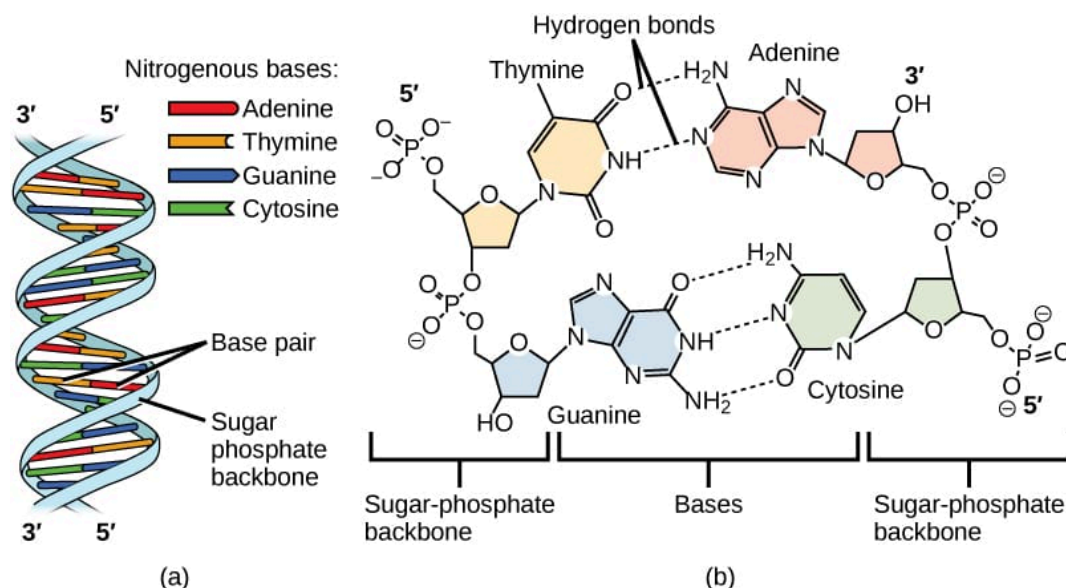


Figure 1.7 DNA (a) forms a double stranded helix, and (b) adenine pairs with thymine and cytosine pairs with guanine. (credit a: modification of work by Jerome Walker, Dennis Myts)

The Structure of RNA

The second nucleic acid in all cells is RNA, or ribonucleic acid. 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. This missing single oxygen atom makes DNA significantly more chemically stable than RNA since oxygen is a "reactive" molecule (Figure 1.8). An easy way to remember the difference is that DNA is "deoxygenated".

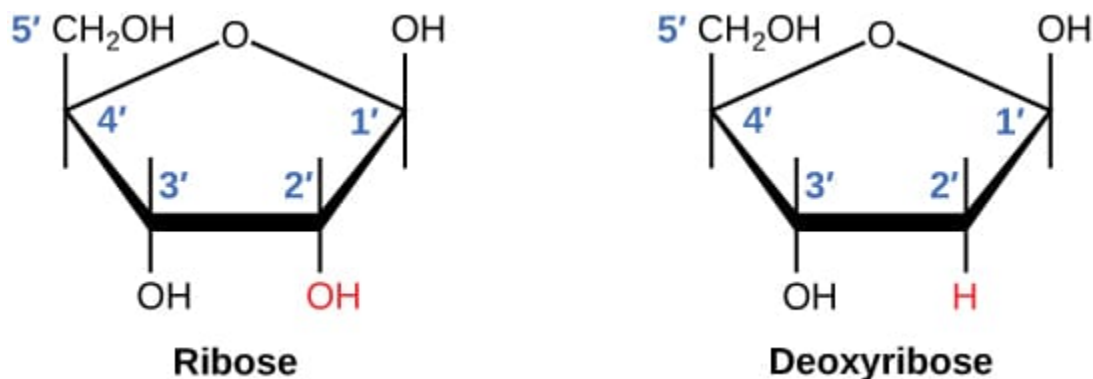


Figure 1.8 The difference between the ribose found in RNA and the deoxyribose found in DNA is that ribose has a hydroxyl group at the 2' carbon.

RNA nucleotides contain the nitrogenous bases adenine, cytosine, and guanine. However, they do not contain thymine, instead RNA uses uracil (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.

Levels of DNA Structure

The structural organization of DNA can be broken down into three distinct hierarchical levels: primary, secondary, and tertiary structure. These levels of organization allow it to function as the genetic material by storing information efficiently, remaining

stable, and fitting inside cells. The **primary structure** of DNA refers to its linear sequence of nucleotides linked together by covalent phosphodiester bonds. This sequence contains the actual genetic information, read in a specific 5' to 3' direction, and consists of a unique order of the four nitrogenous bases.

The **secondary structure** emerges when two of these primary polynucleotide strands run antiparallel to one another and twist into the double helix. The double helix shape is stabilized by hydrogen bonding between complementary base pairs as well as hydrophobic interactions known as base-stacking. The secondary structure protects and stabilizes the genetic information. Under the normal physiological conditions of a living cell, DNA takes on its most stable structural form, known as **B-DNA**. This is the classic, right-handed helix characterized by a uniform diameter and approximately 10.5 base pairs per complete twist. As the two backbones wrap around each other, they leave uneven gaps between the strands, the wide **major groove** and the narrow **minor groove**.

Finally, the **tertiary structure** represents the higher-order arrangement and packaging of the double helix within the cell. A cell must fit a large amount of DNA into a microscopic space, and the tertiary structure helps pack it into the cell while keeping it accessible for replication and transcription.

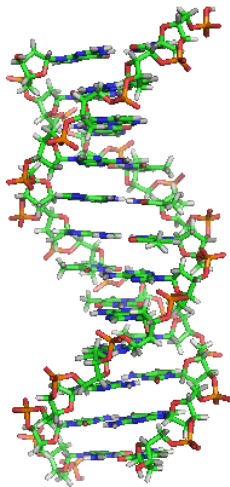


Figure 1.9 The secondary structure of DNA is the double helix.

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 the cell. The DNA of bacteria and eukaryotes are packaged in different ways because of the fundamental structures of the two different cell types. Most prokaryotes contain a *single, circular* chromosome that is found in an area in the cytoplasm called the **nucleoid**.

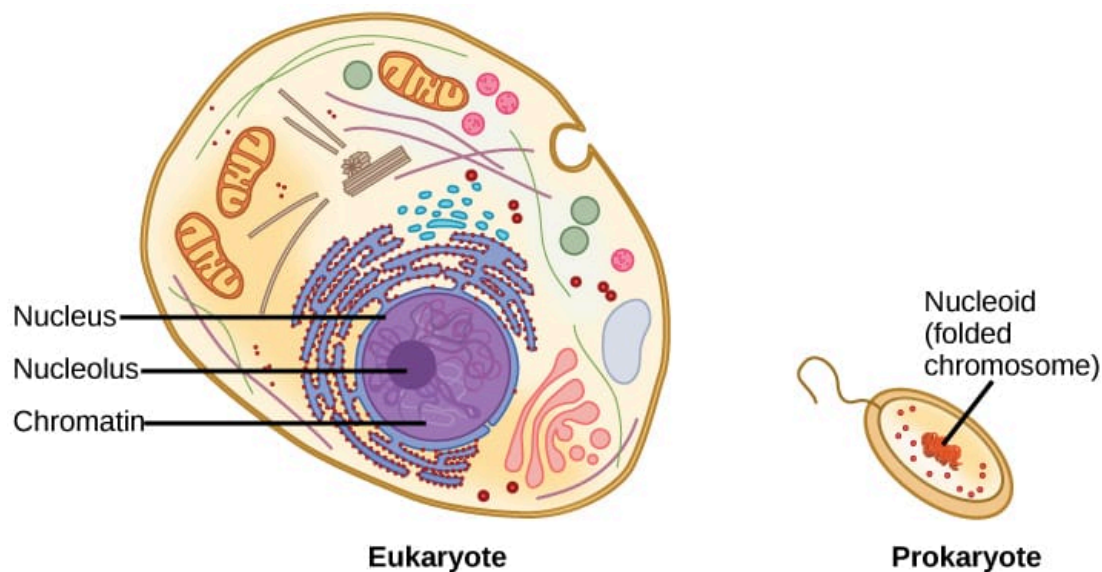


Figure 1.10 A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

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**.

Supercoiling happens when you force extra twists into the helix or deliberately untwist it. **Positive supercoiling** occurs when turns are added to the helix, causing the DNA to twist in the same right-handed direction as the double helix. Because more twists are added into the same length of DNA, the number of base pairs per turn is less than in the relaxed state, resulting in fewer than 10 bp per turn. This over-twisting makes the DNA coil and packs the DNA tightly, but it makes it difficult for the cell to separate the two strands, which can bring replication and transcription to a stop. **Negative supercoiling** occurs when turns are removed, under-winding the helix in the opposite direction. Because the twists are being "loosened", the remaining turns are stretched out resulting in more than 10 base pairs per turn. This is the predominant form of supercoiling found in living cells. Negative supercoiling provides a double advantage, it compacts the DNA into a tiny space, and because it is under-wound, it creates built-in physical tension that actually helps the cell easily unzip the two strands whenever it needs to read the code.

To compress their chromosome, bacteria use active supercoiling with structural looping. First, a type of topoisomerase enzyme called **DNA gyrase** uses ATP to introduce negative supercoils that twist the circular chromosome tightly upon itself, reducing its physical volume while creating localized tension that assists replication and transcription. Second, **nucleoid-associated proteins (NAPs)**—such as HU, IHF, H-NS, Fis, and Dps—bind along the chromosome to bend, bridge, and organize the DNA into massive loop domains. Unlike the proteins eukaryotes use, bacterial NAPs do not form rigid barrels for wrapping; instead, they dynamically alter the DNA's path to simultaneously compact the genome, regulate gene expression, and assist in chromosome segregation during cell division.

In contrast, evolutionary differences emerge within the domain Archaea. While most archaeal species also possess circular chromosomes located in a nucleoid, many organize their genomes using true histone proteins that closely resemble those found in eukaryotes. Instead of relying solely on bacterial-style NAPs, these archaeal histones form dimer or tetramer complexes that bind directly to the DNA double helix, wrapping it into compact structural cores. Consequently, while chromosome packaging varies among diverse archaeal groups, histone-bearing species achieve compaction through a combination of active DNA supercoiling and histone wrapping complexes.

Unlike prokaryotes, eukaryotic cells must pack their *linear* DNA molecules inside a membrane-bound nucleus. To achieve this, the cell combines DNA with proteins to form a complex known as **chromatin** (Figure 1.11). At the most basic level, DNA is wrapped around proteins known as **histones** to form structures called **nucleosomes**. The histone is an octamer assembled from two copies each of four distinct, highly conserved core proteins: **H2A**, **H2B**, **H3**, and **H4**. Histone proteins are rich in amino acids (like lysine and arginine) that carry a positive charge. The negatively charged DNA backbone locks tightly onto the positive histone core.

Connecting one nucleosome bead to the next is a short stretch of exposed DNA called **linker DNA**, which typically ranges from

20 to 80 base pairs in length depending on the species and cell type. Together, the 146 base pairs of core wrapped DNA plus the adjacent linker DNA combine to form a total repeating structural unit of roughly **200 base pairs** per nucleosome. Sitting directly on this linker region is a fifth, distinct protein known as the **H1 histone** (or the linker histone). The H1 protein acts like a clip, locking the wrapped DNA securely to the histone core and pulling neighboring nucleosomes closer together into a tighter fiber. 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.”

A chain of single nucleosome beads compacts a DNA strand significantly, reducing its length by about six-fold into a fiber that measures 10 nanometer (nm) in width. However, this 10-nm “beads-on-a-string” fiber is still far too long to fit inside a cell's nucleus. The cell must continue folding this fiber into increasingly dense layers. With the help of H1 linker histones, the chain of nucleosomes twists and coils tightly upon itself, spinning into a thick molecular rope measuring 30 nm in diameter, compacting the DNA an additional six-fold. Next, the cell takes this 30-nm rope and folds it back and forth into massive, looping arches. The bases of these loops are anchored securely to a central, non-histone structural frame known as the chromosome scaffold, built from specialized **scaffolding proteins**. This looping expands the width of the fiber to 300 nm and keeps distinct regions of the genome organized so they do not tangle.

When a cell prepares for division, it triggers maximum compaction. The scaffold loops twist and compress even further, packing the fiber into tightly condensed chromatids measuring 700 nm across. When duplicated, the iconic X-shaped mitotic chromosome spans 1400 nm in width. At this level, the original two-meter string of human DNA has been compressed more than 10,000-fold.

Euchromatin vs. Heterochromatin

In interphase, the phase of the cell cycle between mitoses at which the chromosomes are de-condensed, eukaryotic chromosomes have two distinct regions that can be distinguished (Figure 1.12). There is a tightly packaged region and a less dense region. **Euchromatin** represents the less dense, accessible regions of the chromosome that typically exist in the 10-nm to 30-nm fiber states. Because the nucleosomes are spaced relatively far apart, the underlying DNA sequence is physically exposed to the cellular environment, allowing transcription enzymes to easily dock onto the DNA template and transcribe genes. Consequently, euchromatin is *transcriptionally active* and contains the vast majority of the genes a cell actively uses.

Heterochromatin is the dense, tightly compacted regions of the chromosome that are packed into the 300-nm loop stage or higher. In this structural state, the nucleosomes are jammed tightly together, blocking access to the DNA. Because transcription machinery cannot physically squeeze between the tightly packed histones, heterochromatin is *transcriptionally silent*. Cells use heterochromatin in two distinct ways. **Constitutive Heterochromatin** is regions of the chromosome that remain permanently locked down across all cell types and throughout the organism's entire life. This is usually found in the centromere (the attachment point for cell division) and the telomeres (the protective caps at the ends of linear chromosomes). These structural regions contain very few active genes. **Facultative Heterochromatin**, on the other hand, is regions of the chromosome where the DNA is dynamically packed or unpack depending on its immediate needs or specific cell type. For example, a liver cell will pack skin-specific genes into silent heterochromatin, while a skin cell will keep those exact same regions open as active euchromatin.

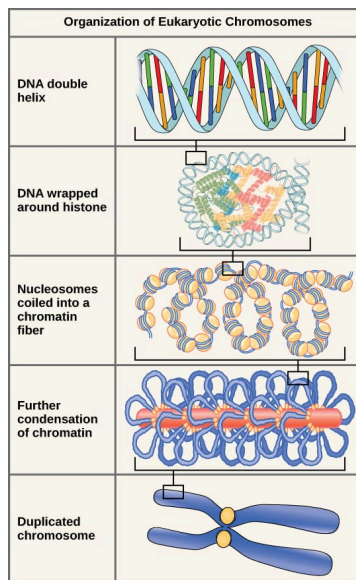


Figure 1.11 These figures illustrate the compaction of the eukaryotic chromosome.

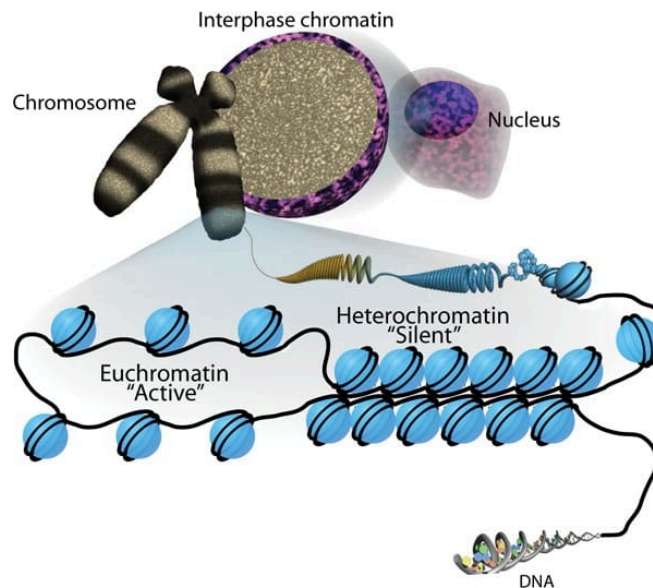
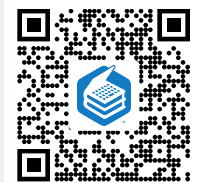


Figure 1.12 In euchromatin the nucleosomes are packed loosely, allowing transcription machinery to easily get to the DNA. In heterochromatin, the nucleosomes are tightly packed, silencing genes.

CONCEPT IN ACTION

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Watch this [animation](#) of DNA packaging.

Key Terms

Primary structure

The linear sequence of nucleotides in DNA.

Secondary structure

The double-helical arrangement of two DNA strands.

Tertiary structure

Higher-order folding and packaging of DNA.

Phosphodiester bond

The covalent bond that links neighboring nucleotides.

Antiparallel

The opposite orientation of DNA strands.

Complementary base pairing

Specific pairing of A with T and G with C.

B-DNA

The most common form of DNA found in cells.

Nucleoid

The region of a bacterial cell containing the chromosome.

Chromatin

The DNA-protein complex that forms eukaryotic chromosomes.

Supercoiling

The twisting of DNA upon itself to create a more compact structure.

Relaxed DNA

DNA that possesses the standard, baseline number of helical turns (~10 to 10.5 base pairs per turn).

Negative Supercoiling

The under-winding of the DNA double helix that compacts the molecule and makes the strands easier to unzip and read.

Positive Supercoiling

The over-winding of the DNA double helix that tightly packs the molecule but makes strand separation extremely difficult.

DNA Gyrase

A specialized topoisomerase found in bacteria that actively consumes energy (ATP) to inject negative supercoils into DNA for structural packing.

Nucleoid-Associated Proteins (NAPs)

A group of bacterial DNA-binding proteins that organize, compact, and regulate the chromosome.

Histone

A positively charged protein around which DNA wraps.

Nucleosome

The fundamental repeating unit of chromatin consisting of DNA wrapped around a histone octamer.

Histone Octamer

A complex containing two copies each of H2A, H2B, H3, and H4.

Linker DNA

The stretch of DNA connecting neighboring nucleosomes.

Linker Histone (H1)

The distinct histone protein that binds directly to the linker DNA, acting like a structural clip to lock the nucleosome in place and draw neighboring beads together.

Euchromatin

Less condensed chromatin that is often transcriptionally active.

Heterochromatin

Highly condensed chromatin that is usually transcriptionally inactive.

Check Your Understanding

1. What are the three parts of a nucleotide?
2. What are the differences between RNA and DNA?
3. What is meant by the primary structure of DNA?
4. Why are phosphodiester bonds important?
5. What does antiparallel mean?
6. Which bases pair together in DNA?
7. Why are G-C rich regions more stable than A-T rich regions?
8. What information can be determined using Chargaff's Rule?
9. What is the difference between secondary and tertiary DNA structure?
10. Why is DNA packaging necessary?
11. Why must DNA be both compact and accessible?
12. How do bacteria and eukaryotes differ in chromosome organization?
13. What is the nucleoid, and how does it differ from the eukaryotic nucleus?
14. What is supercoiling?
15. What are nucleoid-associated proteins, and what roles do they play in bacterial cells?
16. Why do histones bind strongly to DNA?
17. What proteins make up a histone octamer?
18. Approximately how many base pairs are wrapped around a nucleosome core?
19. How does euchromatin differ from heterochromatin?
20. Why is chromatin considered a dynamic structure?

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [1.1.2: The Structure of DNA](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [9.1: The Structure of DNA](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

1.1.3: Genome Organization

Genomic DNA

Before discussing the steps a cell must undertake to replicate and divide its DNA (Chapter 2), a deeper understanding of the structure and function of a cell's genetic material is necessary. A **genome** is the complete set of genetic material present in an organism. Every somatic cell in an organism generally contains the exact same genome, though different sets of genes are selectively activated depending on the specific cell type.

In prokaryotes, the genome is composed of a single, double-stranded DNA molecule in the form of a loop or circle. Up to 90% of a prokaryotic genome consists of coding DNA, or **genes**, meaning the sequences directly translate into proteins or functional RNA molecules. Genes are packed tightly back-to-back, with almost no space between them. Some prokaryotes also have smaller loops of DNA called **plasmids** that are not essential for normal growth, but contain accessory genes. Prokaryotes can exchange these plasmids with other prokaryotes, in a process called **conjugation**. Sometimes a cell receives beneficial new genes that it can add to their chromosomal DNA. *Antibiotic resistance* is one trait that often spreads through a bacterial colony through plasmid exchange from resistant donors to recipient cells.

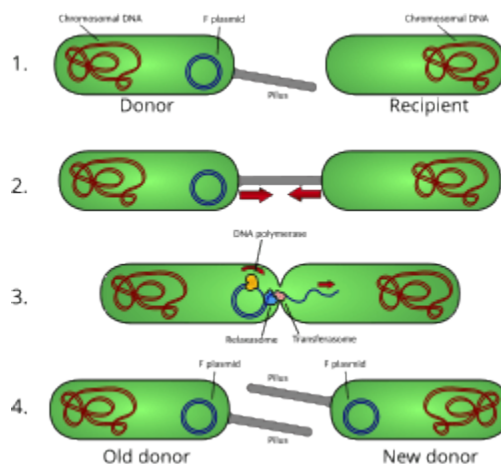


Figure 1.14 Prokaryotes can share plasmids during the process of conjugation.

In eukaryotes, the genome consists of multiple double-stranded linear DNA molecules (Figure 1.15). Each species of eukaryotes has a characteristic number of chromosomes in the nuclei of its cells. Human body (somatic) cells have 46 chromosomes, while human gametes (sperm or eggs) have 23 chromosomes each. A typical body cell contains two matched or homologous sets of chromosomes (one set from each biological parent)—a configuration known as **diploid**. (Note: The letter n is used to represent a single set of chromosomes; therefore, a diploid organism is designated $2n$.) Human cells that contain one set of chromosomes are called gametes, or sex cells; these are eggs and sperm, and are designated $1n$, or **haploid**.

In humans, less than 2% of the genome actually codes for proteins. The remaining 98% is comprised of **non-coding sequences**, or long stretches of repeating sequences, regulatory switches, and structural regions interspersed between widely scattered genes.

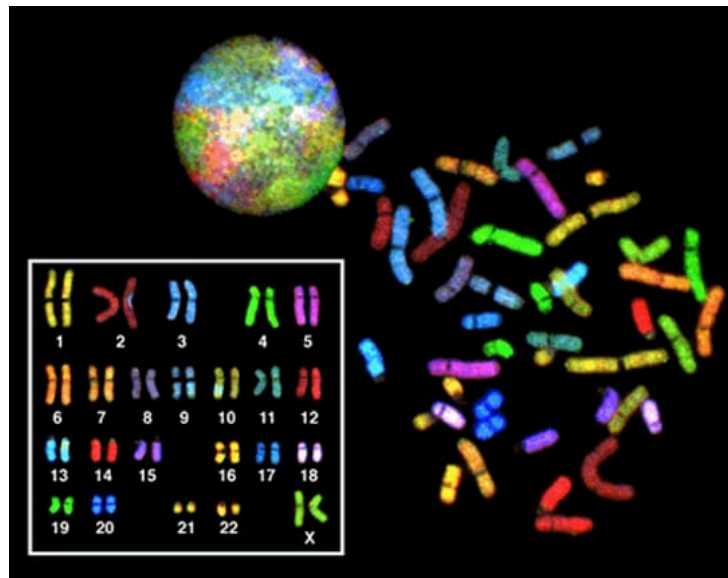


Figure 1.15 There are 23 pairs of homologous chromosomes in a female human somatic cell. The condensed chromosomes are viewed within the nucleus (top), removed from a cell during mitosis (also called karyokinesis or nuclear division) and spread out on a slide (right), and artificially arranged according to length (left); an arrangement like this is called a karyotype. In this image, the chromosomes were exposed to fluorescent stains for differentiation of the different chromosomes. A method of staining called “*chromosome painting*” employs fluorescent dyes that highlight chromosomes in different colors. (credit: National Human Genome Project/NIH)

Genome Size and the C-Value Paradox

Genome size varies among living organisms. Reviewing the approximate *haploid* genome sizes below reveals a surprising trend:

- Bacteriophage: 50,000 bp
- Escherichia coli: 4.6 million bp
- Saccharomyces cerevisiae (Yeast): 12 million bp
- Drosophila melanogaster (Fruit Fly): 170 million bp
- Homo sapiens (Human): 3.2 billion bp
- Amphiuma (Salamander): 85 billion bp

Notice that the salamander genome is ~30 times larger than the human genome. Does this mean salamanders are 30 times more complex than humans? This classic puzzle is known as the **C-value paradox**, where the 'C-value' is the constant amount of DNA found within a species' haploid genome. The paradox describes the complete lack of correlation between an organism's physical complexity and its total genome size. Instead of containing more functional, protein-coding genes, massive genomes like that of the salamander are largely inflated by expanses of non-coding DNA (primarily repetitive sequences and transposable elements) that accumulate over evolutionary time without being cleared out.

The Three Classes of Eukaryotic Genomic DNA

Molecular biologists group DNA sequences into three major categories based on how many times a specific sequence appears in genome. **Unique-sequence DNA** consists of nucleotide sequences that appear only once, or at most a few times, within the entire genome. This category contains the classic protein-coding genes that dictate an organism's physical traits. While unique-sequence DNA contains genes that exist as single copies, evolution frequently duplicates a gene and modifies it over millions of years. This process creates a **gene family**, a cluster of closely related genes that share a similar nucleotide sequence and protein structure, but have evolved to perform slightly different, specialized tasks. A classic example is the globin gene family, which encodes the subunits responsible for carrying oxygen in your blood. During development, embryonic and fetal globins are turned on in the womb. These specialized proteins have an exceptionally high binding affinity for oxygen, allowing the fetus to successfully extract oxygen from the mother's bloodstream. Around the time of birth, the cell flips a genetic switch, silencing the fetal genes and activating adult globin genes, which are optimized for extracting oxygen via the lungs.

Moderately repetitive sequences consist of sequences ranging from a few hundred to several thousand base pairs long that repeat thousands of times across the genome. The cell utilizes this category in two distinct arrangements. In **tandem repeats**, copies occur immediately adjacent to one another in a back-to-back cluster (e.g., 5' -ATG-ATG-ATG-ATG-3'). The cell uses tandem repeats for functional multi-copies. Some genes are needed in such massive quantities that a single copy cannot keep up with cellular demand. The cell maintains hundreds of repeating tandem copies of genes for ribosomal RNA (rRNA) and transfer RNA (tRNA) to ensure protein synthesis never runs out of parts. In **interspersed repeats**, copies are not clumped together in one spot, but are instead scattered broadly across completely different chromosomes throughout the entire genome, separated by millions of base pairs of unique-sequence DNA. The majority of interspersed DNA is composed of mobile, self-replicating molecular hitchhikers known as **transposable elements**. Interspersed repeats make up approximately half of the human genome and are grouped into two primary classes, SINEs and LINEs. **SINEs (short interspersed nuclear elements)** are short repetitive elements, typically fewer than 500 base pairs long. The most common human example is the Alu element that influences gene expression. The Alu sequence has copied itself over a million times, accounting alone for an astonishing 11% of the entire human genome. **LINEs (long interspersed nuclear elements)** are much larger repetitive sequences, often stretching up to several thousand base pairs in length. Unlike smaller SINEs, LINEs carry the active molecular machinery (genes encoding functional proteins) required to physically copy themselves and leap into completely new genomic locations.

Highly repetitive DNA consists of short sequences (often fewer than 10 base pairs long) that repeat in back-to-back clusters hundreds of thousands of times. These are often called **satellite DNA** because their unique nucleotide ratios give them a distinct density profile that separates from the bulk of genomic DNA during laboratory density-gradient centrifugation. Highly repetitive DNA rarely contains actual genetic instructions for proteins. Instead, it serves as structural reinforcement. It forms the **centromeres** (the chromosomal regions where spindle fibers attach to drag chromosomes during cell division) and the **telomeres** (the protective caps on the ends of linear chromosomes).

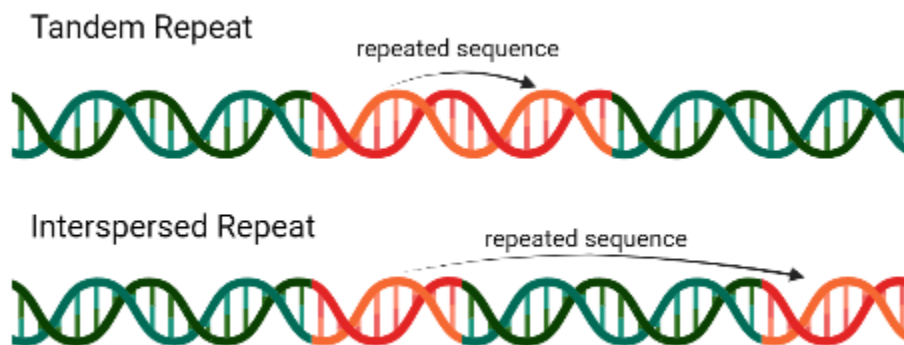


Figure 1.16 Tandem repeats vs. interspersed repeats in the genome.

Jumping Genes

One of the most disruptive discoveries in genetics was made by geneticist Barbara McClintock in the 1940's. While studying the variegation (changing colors) of corn kernels, McClintock discovered that the genome is not fixed. Instead, it contains mobile DNA sequences capable of physically moving from one chromosomal location to another. She named them **transposable elements (TEs)**, commonly known today as "jumping genes." Retrotransposons move using an RNA intermediate, behaving like a "copy and paste". The original transposon sequence in the DNA is transcribed into an RNA molecule. A specialized enzyme called reverse transcriptase reads that RNA copy and converts it back into a new double-stranded DNA fragment. This new DNA duplicate is then inserted into a completely new location in the genome. Because the original copy stays fully intact at its original site, retrotransposons increase the overall size of a genome over evolutionary time. Human *Alu* and *LINE-1* elements move via this mechanism. DNA transposons do not use an RNA intermediate. Instead, they rely on a single specialized enzyme called transposase to move in a "cut and paste" way. The transposase enzyme binds to both ends of the transposon segment, physically clips the entire DNA sequence completely out of its original home, carries that loose DNA segment across the nucleus, and splices it directly into a new target location.

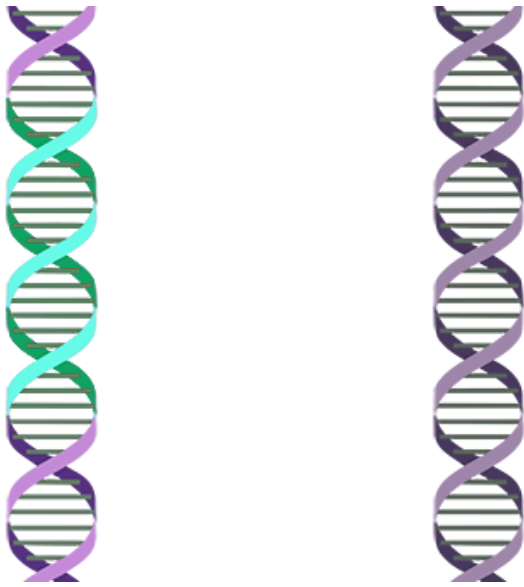


Figure 1.17 "Cut and paste" mechanism of transposition. Scissors indicate transposase. Transposase makes a double-strand DNA break on either side of a transposon, cutting it. Then another double-strand break is made in a target DNA, pasting transposon there. Transposase could be encoded in this transposon or another one.

Eukaryotic Chromosomal Structure and Compaction

If the DNA from all 46 chromosomes in a human cell nucleus were laid out end-to-end, it would measure approximately two meters; however, its diameter would be only 2 nm! Considering that the size of a typical human cell is about 10 μm (100,000 cells lined up to equal one meter), DNA must be *tightly packaged* to fit in the cell's nucleus. At the same time, it must also be readily accessible for the genes to be expressed. For this reason, the long strands of DNA are condensed into compact chromosomes during certain stages of the cell cycle. There are a number of ways that chromosomes are compacted.

In the first level of compaction, short stretches of the DNA double helix wrap around a core of eight histone proteins at regular intervals along the entire length of the chromosome (Figure 10.4). The DNA-histone complex is called chromatin. The beadlike, histone DNA complex is called a nucleosome, and DNA connecting the nucleosomes is called linker DNA. A DNA molecule in this form is about seven times shorter than the double helix without the histones, and the beads are about 10 nm in diameter, in contrast with the 2-nm diameter of a DNA double helix.

The second level of compaction occurs as the nucleosomes and the linker DNA between them coil into a 30-nm chromatin fiber. This coiling further *condenses* the chromosome so that it is now about 50 times shorter than the extended form.

In the third level of compaction, a variety of *fibrous proteins* is used to "pack the chromatin." These fibrous proteins also ensure that each chromosome in a non-dividing cell occupies a particular area of the nucleus that does not overlap with that of any other chromosome (see the top image in Figure 10.3).

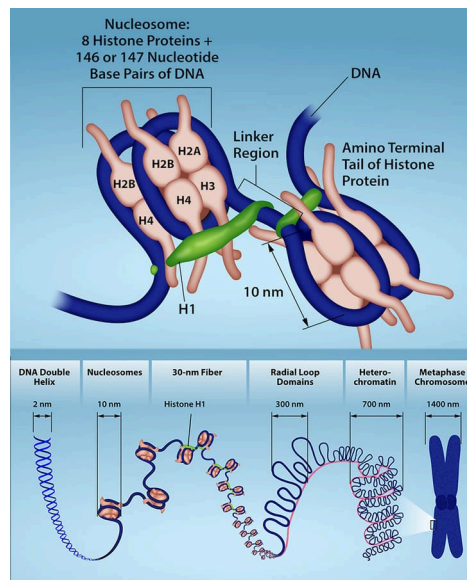


Figure 10.4 Each linear chromosome in a eukaryotic cell is packaged into chromatin, a combination of DNA and proteins. The double-stranded DNA helix associates with the core histones to form nucleosomes. These nucleosomes are further organized into a 30 nm fiber by the linker histone, H1. The fiber then associates with additional proteins to form loops and higher-order heterochromatin packing. DNA packing reaches its most condensed state during metaphase in mitosis in preparation for chromosome separation. Chromatin packing is dynamic and undergoes reversible changes in response to changes in gene expression and the cell cycle. Credit: Rao, A., Ryan, K. Fletcher, S. Hawkins, A. and Tag, A. Department of Biology, Texas A&M University.

DNA replicates in the S phase of interphase, which technically is not a part of mitosis, but must always precede it. After replication, the chromosomes are composed of two linked sister chromatids. When fully compact, the pairs of identically packed chromosomes are bound to each other by cohesin proteins. The connection between the sister chromatids is closest in a region called the centromere. The conjoined sister chromatids, with a diameter of about 1 μm , are visible under a light microscope.

Key Terms

Genome

The complete set of genetic material in an organism.

Gene

Segment of DNA that has the instructions to make a protein or RNA molecule.

Diploid

A cell containing two complete sets of chromosomes (2n).

Haploid

A cell containing a single set of chromosomes (n).

C-value

The amount of DNA in a haploid genome.

C-value paradox

The observation that genome size does not correlate directly with organismal complexity.

Unique-sequence DNA

DNA present in only one or a few copies within a genome.

Gene family

A group of related genes that arose through duplication of an ancestral gene.

Moderately repetitive DNA

DNA sequences present in hundreds to thousands of copies.

Highly repetitive DNA

DNA sequences present in thousands to millions of copies.

SINE

Short interspersed nuclear element.

LINE

Long interspersed nuclear element.

Check Your Understanding

1. What is a genome?
2. What is the C-value paradox?
3. Why does genome size not necessarily reflect organismal complexity?
4. What is the difference between unique-sequence DNA and repetitive DNA?
5. How do gene families arise?
6. Why do cells contain multiple copies of rRNA and tRNA genes?
7. What are SINEs and LINEs?
8. Why is repetitive DNA important even when it does not encode proteins?

1.2: DNA REPLICATION

1.2.1: Basics of DNA Replication

1.2.2: DNA Replication

1.2.3: DNA Replication in Bacteria

1.2.4: DNA Replication in Eukaryotes

1.2.1: Basics of DNA Replication

Learning Objectives

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

- Explain how the structure of DNA reveals the replication process
- Describe the Meselson and Stahl experiments

The elucidation of the structure of the double helix provided a hint as to how DNA divides and makes copies of itself. In their 1953 paper, Watson and Crick penned an incredible understatement: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." With specific base pairs, the sequence of one DNA strand can be predicted from its complement. The double-helix model 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. What was not clear was how the replication took place. There were three models suggested (Figure 14.12): *conservative*, *semi-conservative*, and *dispersive*.

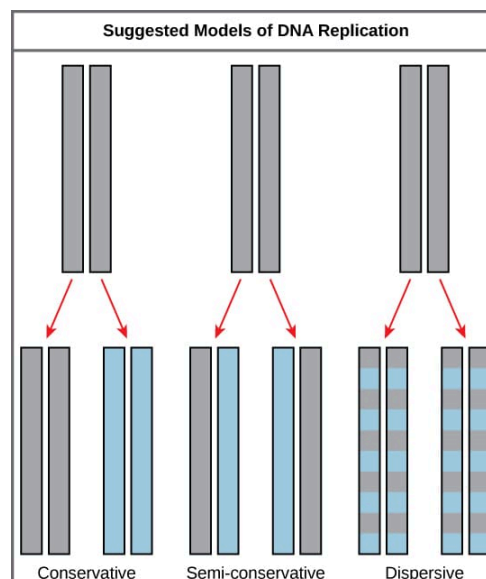


Figure 14.12 The three suggested models of DNA replication. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

In conservative replication, the parental DNA remains together, and the newly formed daughter strands are together. The semi-conservative method suggests that each of the two parental DNA strands acts as a template for new DNA to be synthesized; after replication, each double-stranded DNA includes one parental or "old" strand and one "new" strand. In the dispersive model, both copies of DNA have double-stranded segments of parental DNA and newly synthesized DNA interspersed.

Meselson and Stahl were interested in understanding how DNA replicates. They grew *E. coli* for several generations in a medium containing a "heavy" isotope of nitrogen (^{15}N), which gets incorporated into nitrogenous bases, and eventually into the DNA (Figure 14.13).

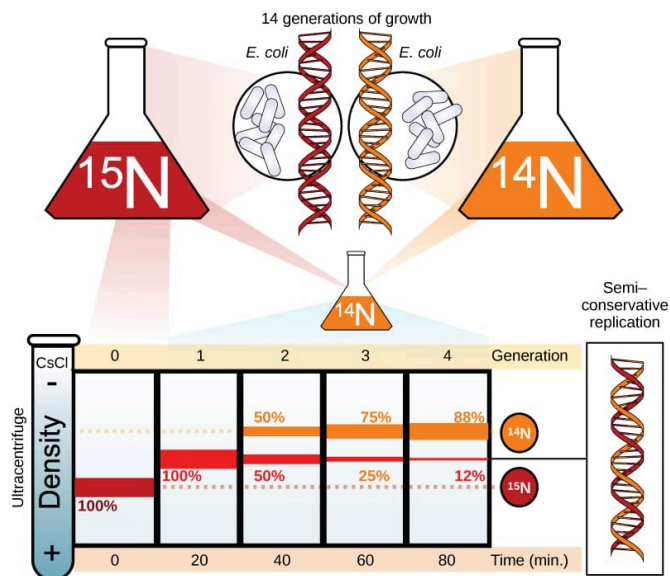


Figure 14.13 Meselson and Stahl experimented with *E. coli* grown first in heavy nitrogen (^{15}N) then in ^{14}N . DNA grown in ^{15}N (red band) is heavier than DNA grown in ^{14}N (orange band), and sediments to a lower level in cesium chloride solution in an ultracentrifuge. When DNA grown in ^{15}N is switched to media containing ^{14}N , after one round of cell division the DNA sediments halfway between the ^{15}N and ^{14}N levels, indicating that it now contains fifty percent ^{14}N . In subsequent cell divisions, an increasing amount of DNA contains ^{14}N only. These data support the semi-conservative replication model. (credit: modification of work by Mariana Ruiz Villareal)

The *E. coli* culture was then placed into medium containing ^{14}N and allowed to grow for several generations. After each of the first few generations, the cells were harvested and the DNA was isolated, then centrifuged at high speeds in an ultracentrifuge. During the centrifugation, the DNA was loaded into a *gradient* (typically a solution of salt such as cesium chloride or sucrose) and spun at high speeds of 50,000 to 60,000 rpm. Under these circumstances, the DNA will form a band according to its *buoyant density*: the density within the gradient at which it floats. DNA grown in ^{15}N will form a band at a higher density position (i.e., farther down the centrifuge tube) than that grown in ^{14}N . Meselson and Stahl noted that after one generation of growth in ^{14}N after they had been shifted from ^{15}N , the single band observed was intermediate in position in between DNA of cells grown exclusively in ^{15}N and ^{14}N . This suggested either a semi-conservative or dispersive mode of replication. The DNA harvested from cells grown for two generations in ^{14}N formed two bands: one DNA band was at the intermediate position between ^{15}N and ^{14}N , and the other corresponded to the band of ^{14}N DNA. These results could only be explained if DNA replicates in a semi-conservative manner. And for this reason, therefore, the other two models were ruled out.

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 strands will be complementary to the parental or “old” strands. When two daughter DNA copies are formed, they have the same sequence and are divided equally into the two daughter cells.

[Link to Learning](#)

Link to Learning

View [this video](#) on DNA replication.

1.2.2: DNA Replication

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. 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 3' AGTCATGA 5' will have a complementary strand with the sequence 5' TCAGTACT 3' (Figure 2.3).

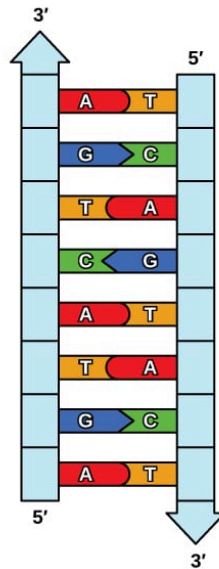


Figure 2.3 The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

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 2.4). 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.

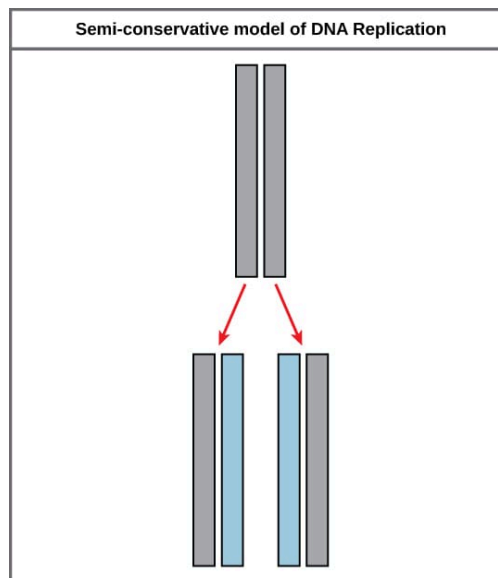


Figure 2.4: The semiconservative model of DNA replication is shown. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

The Biochemistry of Replication

To make a new strand of DNA, the replication machinery requires both physical material and a continuous input of chemical energy. Evolution solved this through the use of deoxyribonucleoside triphosphates (dNTPs), substrates that package a monomeric nucleotide building block and thermodynamic energy within the same molecule. Each dNTP consists of a nitrogenous base, a central five-carbon deoxyribose sugar, and a highly unstable *triphosphate* tail containing high-energy phosphoanhydride bonds. Because the negatively charged phosphate groups naturally repel one another like compressed springs, the bonds hold immense potential energy. During DNA synthesis, the enzyme **DNA polymerase** coordinates magnesium ions within its active site to lower the pK_a of the growing strand's terminal hydroxyl ($3'-OH$) group, rendering it highly nucleophilic which then launches a nucleophilic attack on the electrophilic alpha (α) phosphate of the incoming dNTP, cleaving the bond linking it to the beta (β) and gamma (γ) phosphates. This *condensation reaction* permanently attaches the nucleotide monophosphate to the growing sugar-phosphate backbone and expels the remaining phosphates as a pyrophosphate (PP_i) leaving group. Intracellular pyrophosphatases instantly hydrolyze this released pyrophosphate into two independent inorganic phosphates (P_i), which releases a massive amount of free energy (ΔG), shifts the chemical equilibrium, and renders DNA polymerization energetically exergonic and effectively irreversible.

Rather than passively waiting for the correct nucleotide to drift into place, the catalytic active site of DNA polymerase actively enforces complementary base pairing. dNTP selection relies first on the precise configuration of hydrogen bond donors and acceptors, ensuring adenine pairs exclusively with thymine via two stabilizing bonds, and guanine with cytosine via three. Mismatched nucleotides fail to form stable bonds and rapidly dissociate from the active site before a covalent bond can form. Second, the enzyme enforces structural uniformity along the DNA backbone. The active site of DNA polymerase is a tight, rigid pocket designed to accommodate a width of exactly $10.8 \text{ \AA}$, which is the precise spatial dimension of a purine-pyrimidine pair. Forcing two large purines together creates a pair too wide to fit the catalytic pocket, while pairing two small pyrimidines leaves the bases physically too far apart to form stable hydrogen bonds. Only when a perfectly matched dNTP enters the active site does the DNA polymerase undergo a conformational change, wrapping tightly around the substrate to trigger the nucleophilic attack, restricting initial insertion mistakes to an astonishingly low rate of just one error per 10^4 to 10^5 insertions.

clipboard_e6434af94aac722fb03377b6a431d2ea9.png

Figure 2.5: A deoxyribonucleoside triphosphate showing the alpha phosphate (purple), beta phosphate (orange), and gamma phosphate (green). During nucleotide polymerization reaction, the gamma and beta phosphates leave as pyrophosphate group.

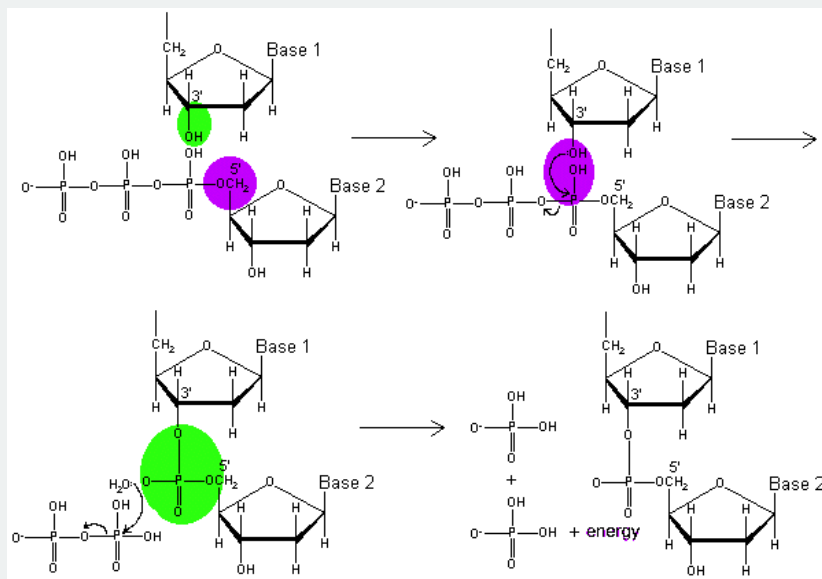


Figure 2.6 The condensation reaction of

phosphodiester bond formation. The 3' hydroxyl group attacks the triphosphate group on the incoming nucleotide. A new phosphodiester bond is formed and a pyrophosphate group leaves.

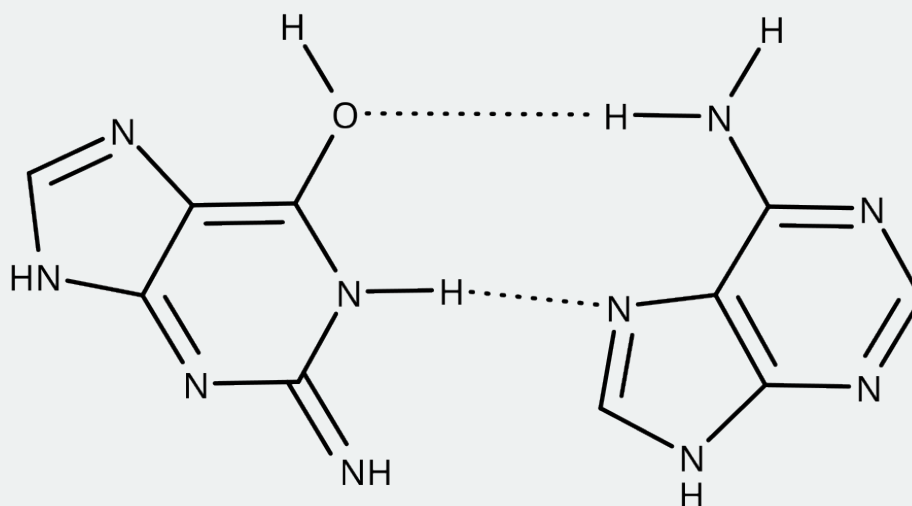


Figure 2.7 Guanine and Adenine pair. The hydrogen bond donors and acceptors fail to align correctly, leaving the bases incapable of forming stable bonds.

A diagram of a moleculeAI-generated content may be incorrect.

Figure 2.8 **Geometric Symmetry of Complementary Base Pairs.** The structural framework of replication fidelity relies on the strict pairing rules of the double helix. Because each pair combines a double-ring purine with a single-ring pyrimidine, they achieve an identical width of **10.8 Å**. This spatial uniformity allows only properly matched base pairs to fit into the tight active site of DNA polymerase, minimizing insertion errors.

DNA Repair

DNA polymerase can make mistakes while adding nucleotides and will occasionally add the wrong one. When this happens, it distorts the helix and stalls the enzyme, it cannot move forward. The DNA strand slips out of the catalytic domain of the polymerase and into the "editing" domain that possesses $3' \rightarrow 5'$ **exonuclease activity**. This exonuclease works backward from the terminal end, snipping out the mismatched nucleotide from the exposed end of the nucleic acid strand. Once the error is removed, the growing strand snaps back into the primary polymerization active site, and normal synthesis resumes. This proofreading capability is the ultimate evolutionary driver behind the directional constraints of replication, explaining why DNA polymerase cannot synthesize DNA *de novo* (from scratch) and must instead extend an existing strand exclusively in the $5' \rightarrow 3'$ direction. Because synthesis relies on a nucleophilic attack originating from the growing strand's $3'$ -OH group toward the incoming $5'$ triphosphate substrate, the universal $5' \rightarrow 3'$ approach ensures that even after a mismatch is excised during proofreading, the growing strand always terminates in a fresh, unspent $3'$ -OH group. If a cell were to synthesize DNA in the reverse $3' \rightarrow 5'$ direction, the high-energy activation energy would reside on the tip of the growing strand; removing a mismatched base would strip away the activating triphosphate group, leaving a chemically inert monophosphate "dead end" incapable of further elongation.

Origins of Replication and Replication Forks

The cell must execute replication across entire chromosomes to prevent incomplete duplication. Replication doesn't begin at random locations, it is initiated at regulated starting points called **origins of replication**, which act as assembly sites for replication proteins and enzymes. To initiate DNA synthesis, regulatory signals activate the origin by recruiting initiator proteins that bind directly to the origin sequence. This binding event twists and distorts the DNA molecule, inducing local strain that forces open the hydrogen bonds holding the two strands together. To minimize the energy required for this initial melting, origin sequences are generally enriched with adenine-thymine (A-T) base pairs; because A-T pairs share only two hydrogen bonds compared to the three found in cytosine-guanine (C-G) pairs. The unwinding creates a single-stranded pocket known as a **replication bubble**, which recruits multi-enzyme complexes called **replisomes**. On the outer edges of the expanding bubble, the Y-shaped junctions where the parental strands actively separate are termed **replication forks**. Virtually all life utilizes *bidirectional* replication, meaning that upon origin activation, two distinct replication forks are established at opposite ends of the bubble, migrating away from the origin in opposite directions.

A diagram of a bubble form AI-generated content may be incorrect.

Figure 2.9 Replication bubbles in DNA. The replication forks are at each end of the bubble where the double helix is actively being unwound.

The segment of DNA replicated from a single origin, including the origin sequence itself and all DNA copied by its corresponding pair of replication forks, is defined as a **replicon**. Because the two strands of a parental DNA molecule are antiparallel, the universal requirement for DNA polymerase to synthesize exclusively in the $5' \rightarrow 3'$ direction creates a challenge at each advancing replication fork. As the fork unwinds in a single direction, the two exposed templates are opposite polarities. The **leading strand** template runs in a $3' \rightarrow 5'$ direction relative to the fork's movement, allowing a DNA polymerase to track *continuously* toward the unwinding junction and synthesize its new partner strand in a uninterrupted, progressive manner. Conversely, the opposing **lagging strand** template is exposed in a $5' \rightarrow 3'$ direction. Because the polymerase must write in the $5' \rightarrow 3'$ direction, it is forced to synthesize *away* from the opening fork. As the moving replisome progressively uncovers new template behind the enzyme, the polymerase must repeatedly jump backward toward the fork, re-initiate synthesis, and build short, *discontinuous* segments known as **Okazaki fragments**.

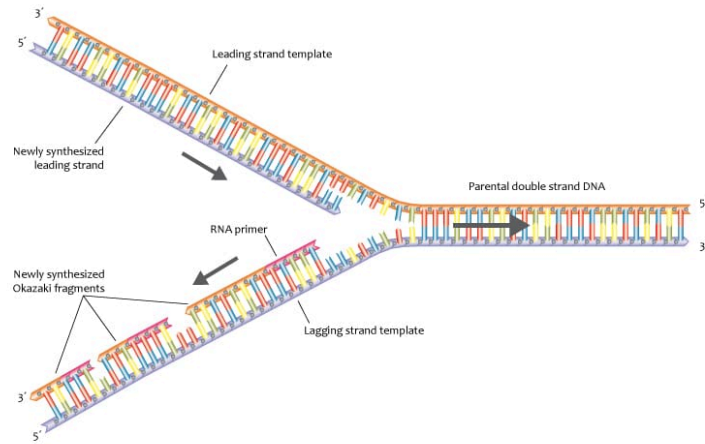


Figure 2.11 Simple schematic of a replication fork showing the synthesis of leading and lagging strands. Pay close attention to the polarity of the template strands and the newly synthesized strands.

The overall number and scale of replicons differ between the major domains of life. Most bacteria have a relatively small, single circular chromosome that functions as a single replicon. In *E. coli*, replication initiates at a single distinct origin sequence called **oriC** and proceeds bidirectionally around the circular molecule. To prevent one fork from overrunning the other, bacterial chromosomes contain a specialized termination zone opposite the origin formed by specific termination (*Ter*) sequences. When bound by *Tus* proteins, this protein-DNA complex acts as a trap, allowing a replication fork to enter the termination zone from one direction but blocking its exit. Once the two opposing forks collide and stall within this trapped zone, the replication complexes disassemble, and specialized topoisomerase enzymes physically disentangle the two interlocking daughter circles. In contrast, eukaryotic chromosomes are linear and larger; if they relied on a single origin, replication would take too long. A single eukaryotic chromosome is organized into thousands of small replicons that fire simultaneously during the cell cycle, each generating its own replication bubble. Eukaryotes lack sequence-specific termination traps; instead, as the bubbles expand outward, neighboring replication forks eventually collide. Once they meet, the replication machinery on both strands disassembles, the remaining gaps in the sugar-phosphate backbone are covalently sealed, and the independent replication events merge into a continuous, perfectly duplicated linear chromosome.

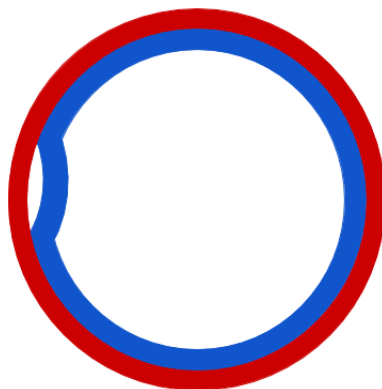


Figure 2.12 DNA replication in bacterial chromosomes. The thick colored lines represent the original parental DNA and the dotted

lines represent the newly synthesized DNA.

The Replisome and Replication Enzymes

To replicate successfully, cells must open a tightly wound helix, prevent the separated strands from rejoining, relieve the torsional tension created during unwinding, add new nucleotides, correct mistakes instantly, and join newly synthesized fragments together. Because no single enzyme possesses the active sites or structural domains to perform all of these tasks, replication is done by a large collection of proteins working simultaneously at the replication forks. These proteins form the replisome.

The initial challenge in DNA replication is separating the two parental strands. While the hydrogen bonds holding complementary nitrogenous bases together are weak individually, they collectively strong across millions of base pairs. This "unzipping" task is performed by **DNA helicase**, which uses the chemical energy stored in adenosine triphosphate (ATP) to physically move down the DNA backbone. As helicase advances, it disrupts the hydrogen bonds between base pairs to separate the parental strands, opening the stable double helix. However, this unwinding action creates an immediate mechanical bottleneck. Much like pulling apart the strands of a tightly twisted rope, separating the strands in the middle compresses the remaining twists into the un-separated sections ahead. This overwinding accumulates as positive supercoiling ahead of the replication fork. Left unchecked, this torsional strain builds into tight, knot-like wraps that would stall the replisome.

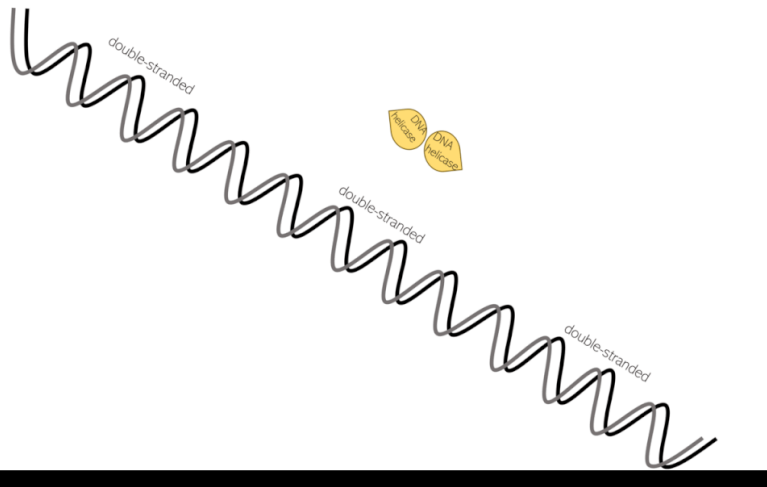


Figure 2.13 Helicase unwinding the DNA.

To resolve this problem, cells use **topoisomerases**, a class of specialized enzymes positioned immediately ahead of the advancing replication fork to regulate DNA topology and provide tension relief. All topoisomerases relieve torsional strain by targeting the overwound DNA, breaking the phosphodiester backbone, allowing the tension to dissipate, and then sealing the backbone. To prevent the broken strands from drifting away or causing chromosome fragmentation, the enzyme has a specialized tyrosine residue in its active site to form a transient, covalent bond with the broken DNA ends. These enzymes are broadly divided into two structural classes based on how many strands they manipulate at one time. **Type I topoisomerases** make a temporary cut in only one of the two strands and function without ATP, relying instead on the built-up tension of the overwound DNA to drive passive relaxation. **Type II topoisomerases** cut both strands of the DNA double helix simultaneously and utilize ATP hydrolysis to actively pass an intact segment of the loop through the double-stranded break.

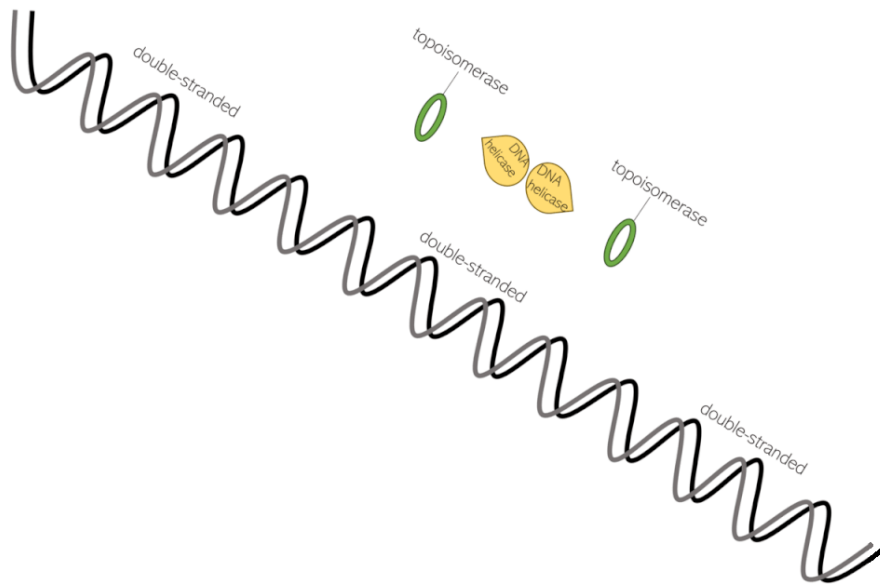


Figure 2.14 Helicase unwinding DNA with topoisomerase relieving supercoiling ahead of the replication fork.

Once topoisomerase and helicase open and relax the double helix, the exposed single-stranded DNA (ssDNA) faces immediate instability. These single strands tend to spontaneously snap back together through hydrogen bonding, or fold backward on themselves to form secondary hairpin structures that physically block the replisome. To prevent that from happening, **single-stranded DNA-binding proteins (SSBs)** bind to the freshly exposed strands. SSBs stabilize the unwound DNA without obscuring the bases, keeping the strands straight, accessible, and protected until the synthesis machinery can pass through.

Once the template strand is exposed and stabilized, DNA polymerases cannot simply begin building a new strand from scratch. These enzymes can only attach a new nucleotide to an existing nucleotide chain that possesses a free 3' hydroxyl ($-OH$) group, they are incapable of binding two free-floating nucleotides together. To overcome this limitation, cells use **primase**, a specialized type of *RNA polymerase* that does not require an existing starter molecule. Primase binds directly to the exposed single-stranded template and synthesizes a short, temporary complementary segment of RNA, typically 5 to 10 nucleotides long, known as an **RNA primer**. Because the final nucleotide of this RNA primer features an exposed, free 3'-OH group, it creates the essential docking site for DNA polymerase to bind and begin synthesis. While the continuously synthesized *leading strand* requires only a single primer at the origin of replication, the discontinuous *lagging strand* requires hundreds or thousands of individual RNA primers, forcing primase to continuously hop ahead and lay down a fresh primer for every segment produced.

The central workhorse of replication is **DNA polymerase**, the enzyme responsible for the elongation of the new DNA strands. Polymerase reads the open template strand, selects a matching deoxynucleoside triphosphate (dNTP) from the surrounding environment, and catalyzes a condensation reaction. This reaction forms a phosphodiester bond between the 5' phosphate group of the incoming nucleotide and the free 3'-OH group of the existing strand. Because of the strict orientation of its active site, DNA polymerase can only extend a growing strand exclusively in the $5' \rightarrow 3'$ direction. Left on its own, a DNA polymerase molecule is prone to falling off its template strand after adding only a few dozen nucleotides. To prevent this, cells use an accessory structural protein called the **sliding clamp**. A specialized multi-protein complex known as the clamp loader uses ATP to open this doughnut-shaped ring and snap it securely around the DNA helix. The sliding clamp encircles the DNA without directly touching the bases, allowing it to move freely along the backbone while binding directly to the rear of the moving DNA polymerase.

Before replication is finished, the temporary RNA primers must be systematically removed and replaced with standard DNA nucleotides by specialized repair polymerases and nucleases. However, there's one final challenge: while enzymes can fill in the missing nucleotides, they cannot join the final 3'-OH group of the newly synthesized piece to the adjacent 5'-phosphate group of the existing fragment ahead of it. This leaves a tiny structural break, or "nick," in the sugar-phosphate backbone. This final sealing task is performed by **DNA ligase**. Using energy derived from ATP or NAD^+ , ligase activates the phosphate backbone to catalyze the final phosphodiester bond, covalently sealing the fragments into a single, continuous, and uninterrupted macromolecule. Ligase is used to connect Okazaki fragments as well.

Interactive Content

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this Interactive Content online](#)



DNA replication video.

Key Terms

Template Strand

A DNA strand used to direct synthesis of a complementary strand.

DNA Polymerase

An enzyme that synthesizes DNA.

dNTP

A deoxyribonucleoside triphosphate used as a substrate for DNA synthesis.

Phosphodiester Bond

The covalent bond that links neighboring nucleotides.

Condensation Reaction

A chemical reaction that covalently bonds two molecules together while eliminating a smaller byproduct molecule.

Exonuclease Activity

The proofreading feature of DNA polymerase that lets it reverse direction, snip out a mismatched nucleotide it just added, and correct its own mistake before moving forward.

5' → 3' Synthesis

The universal direction of DNA synthesis in cells.

Origin of Replication

A DNA sequence where replication begins.

Replication Bubble

A region of unwound DNA where replication is occurring.

Replication Fork

A Y-shaped structure where DNA strands separate and new DNA is synthesized.

Bidirectional Replication

Replication that proceeds in two directions from an origin.

Leading Strand

The newly synthesized DNA strand built continuously in the same direction as the advancing replication fork.

Lagging Strand

The newly synthesized DNA strand built discontinuously in short, backward bursts away from the progression of the replication fork.

Okazaki Fragments

The short, discontinuous segments of DNA synthesized sequentially on the lagging strand template that are later joined together.

Replicon

The DNA region replicated from a single origin.

oriC

The primary origin of replication in *E. coli*.

Helicase

An enzyme that unwinds the DNA double helix.

Topoisomerase

An enzyme that regulates DNA supercoiling and relieves torsional stress.

Single-Stranded Binding Protein (SSB)

A protein that stabilizes separated DNA strands.

Primase

An enzyme that synthesizes RNA primers.

RNA Primer

A short RNA sequence that provides a starting point for DNA synthesis.

Sliding Clamp

A protein that holds DNA polymerase on DNA during replication.

DNA Ligase

An enzyme that seals breaks in the DNA backbone.

Replisome

The protein complex responsible for DNA replication.

Check Your Understanding

1. Why is helicase required during replication?
2. What problem does topoisomerase solve?
3. Why are single-stranded binding proteins necessary?
4. Why does DNA polymerase require an RNA primer?
5. What is the function of the sliding clamp?
6. How does DNA polymerase improve replication accuracy?
7. What role does DNA ligase play?
8. What is the replisome?
9. What is an origin of replication?
10. Why are origins necessary?
11. What is a replication bubble?
12. How many replication forks are present in a replication bubble?
13. What does bidirectional replication mean?
14. Why do eukaryotic chromosomes require multiple origins of replication?
15. What is a replicon?
16. What happens when two replication forks meet?
17. Why does the requirement for 5' → 3' synthesis force the cell to create leading and lagging strand?
18. What role does complementary base pairing play in DNA replication?
19. What is a template strand?
20. What specific chemical bond is formed to link neighboring nucleotides together during synthesis?
21. In which direction is DNA synthesized?
22. Why is DNA polymerase said to read one strand and write another?

This page titled [1.2.2: DNA Replication](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [9.2: DNA Replication](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

1.2.3: DNA Replication in Bacteria

While there are many similarities in the DNA replication between bacteria and eukaryotes, the structural differences in the chromosomes necessitate some differences in the replication process. Bacterial DNA replication is one of the most thoroughly characterized processes in molecular biology, serving as a model for understanding replication in all organisms. Using the model bacterium *Escherichia coli* (*E. coli*) as our primary guide, this section tracks the complete life cycle of bacterial replication.

Replication can be divided into three phases: initiation, elongation, and termination. As stated in the previous section, *E. coli* has a single origin of replication (*oriC*) on its one chromosome, as do most prokaryotes. The *oriC* in *E. coli* is approximately 245 base pairs long and is rich in AT sequences. The transition into the initiation phase happens when the origin is activated. The primary driver of the initiation process is a specialized bacterial protein called **DnaA**. When metabolic conditions dictate that the cell is ready to divide, ATP-bound DnaA proteins locate and bind tightly to the *oriC* sequence. As more DnaA molecules accumulate, they bind to one another, oligomerizing into a tight, core-like spiral protein complex. The template DNA wraps around this DnaA core, introducing physical strain on the adjacent double helix. This forces the nearby, A-T rich region to pop open, creating the replication bubble.

Once the A-T rich region is open, helicase must be loaded onto the single-stranded template. Bacteria use two proteins for this. **DnaB (helicase)** is a ring-shaped motor that uses ATP to unzip the double helix. DnaB is a completely closed ring so it cannot clamp onto an intact strand on its own, so loader protein **DnaC (helicase loader)** is used as well. DnaC binds to DnaB, using ATP to pry the ring open. The DnaC loader then manually guides the open DnaB ring over the exposed single strands within the replication bubble. Two separate DnaB helicase rings are loaded onto the origin in a bidirectional configuration, with one facing clockwise and the other facing counterclockwise. Once the rings snap shut around the DNA, DnaC dissociates, leaving the functional helicases permanently locked onto their respective target strands.

With the helicases successfully loaded, the replication bubble transforms into two independent replication forks pointing in opposite directions. The DnaB helicases begin recruiting the rest of the replication machinery to finalize the assembly of the replisome. **Bacterial primase (DnaG)** attaches directly to the moving helicase ring, preparing to synthesize the first RNA primers. **DNA Polymerase III (DNA pol III)** are loaded onto the strands via their sliding β -clamps, tethering themselves to the helicase complexes. At this point, the initiation complex is fully converted into a pair of replisomes, and the cell officially transitions from initiation into the elongation phase.

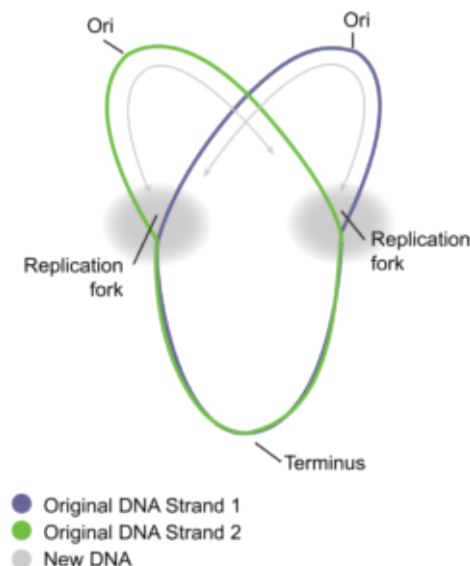


Figure 2.15 DNA replication in prokaryotes.

Once the replication forks break away from the *oriC* origin, the elongation stage begins, where replication occurs simultaneously along both sides of the circular chromosome in bidirectional replication. As the clockwise fork replicates down the right side of the circular chromosome, the counterclockwise fork duplicates the left side at exactly the same speed. Upstream of the synthesis site, the spinning DnaB helicase ring uses ATP to tear through the hydrogen bonds of the double helix, creating positive supercoils

ahead of the fork so **DNA gyrase** (a bacterial Type II topoisomerase) continually rides ahead of the replisome, cutting both DNA strands, untwisting the overwound knots, and sealing the backbone back together.

Single stranded binding proteins coat the exposed single-stranded DNA templates, keeping them straight, accessible, and safe. One DNA Polymerase III tracks directly behind the helicase, continuously adding dNTPs to the 3' end of the leading strand. Concurrently, a second DNA Polymerase III synthesizes the lagging strand backward via the **trombone model** (Replication Fork Video below). In this configuration, primase (DnaG) periodically places an RNA primer near the fork crux, and the polymerase extends it into a 1,000 to 2,000 nucleotide Okazaki fragment before looping out, releasing the template, and resetting for the next cycle.

Because the lagging strand is made of alternating segments of RNA and DNA separated by physical gaps, it must be edited before the replication cycle can conclude. In bacteria, this process is streamlined, relying on DNA Polymerase I and DNA Ligase. When DNA Polymerase III completes an Okazaki fragment, it dissociates when encountering the upstream primer of the previous fragment. **DNA Polymerase I (Pol I)** docks at this single-stranded nick. Using its unique 5' → 3' **exonuclease activity**, Pol I degrades the RNA primer directly ahead of it. Simultaneously, using its internal polymerase domain, it extends the 3' end of the new fragment, replacing every excised RNA nucleotide with a DNA nucleotide. Once Pol I has fully replaced the primer, it dissociates, leaving behind a filled gap but a structural break in the sugar-phosphate backbone. DNA Ligase forms a permanent, covalent phosphodiester bond between the adjacent 3'-OH and 5'-phosphate groups, connecting the fragments into an unbroken, continuous daughter strand. Bacteria also use another polymerase, DNA polymerase II, but it is primarily used for DNA repair.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)

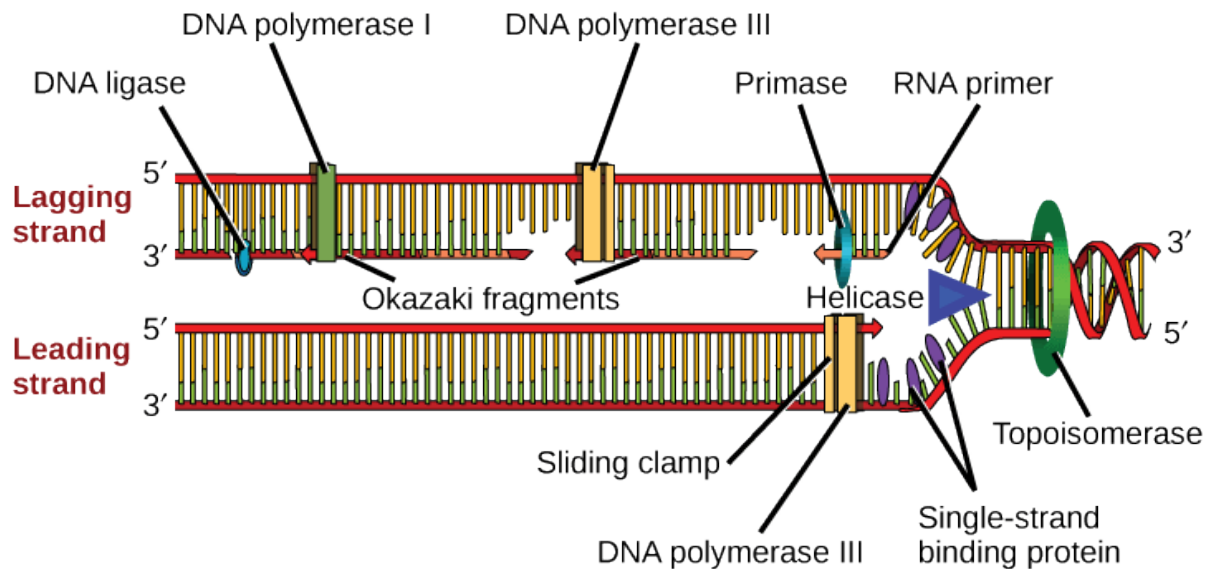


Figure 2.16 A replication fork is formed when helicase separates the DNA strands at the origin of replication. The DNA tends to become more highly coiled ahead of the replication fork. Topoisomerase breaks and reforms DNA's phosphate backbone ahead of the replication fork, thereby relieving the pressure that results from this supercoiling. Single-strand binding proteins bind to the single-stranded DNA to prevent the helix from re-forming. Primase synthesizes an RNA primer. DNA polymerase III uses this primer to synthesize the daughter DNA strand. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches called Okazaki fragments. DNA polymerase I replaces the RNA primer with DNA. DNA ligase seals the gaps between the Okazaki fragments, joining the fragments into a single DNA molecule. (credit: modification of work by Mariana Ruiz Villareal)

Because bacterial chromosomes are closed circles, the two replication forks cannot simply run off the structural ends of the

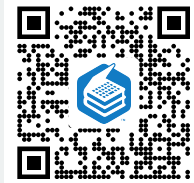
DNA. Instead, they must meet safely on the opposite side of the chromosome to cleanly complete genome duplication without damaging the molecule. If the clockwise and counterclockwise forks traveled at random speeds, they might collide unpredictably, potentially causing the replication machinery to stall, collapse, or break the chromosome. To prevent this, *E. coli* utilizes a specialized trap system consisting of **Ter sites** and **Tus proteins**. *Ter* sites are a cluster of short, directional DNA sequences (labeled *TerA* through *TerJ*) located roughly opposite to the *oriC* origin, while the **Tus protein** is a specialized replication termination protein that binds to these sequences.

The Tus-*Ter* complex acts as a one-way gate for advancing machinery. The complex has a permissive face and a non-permissive face. If an advancing replication fork arrives tracking in the proper direction, it hits the permissive face of the Tus-*Ter* complex; this orientation allows the fork to smoothly pass through, displacing the Tus protein and continuing its synthesis. However, if a fork arrives from the opposite direction, it hits the non-permissive face and induces a conformational change that locks the Tus protein tightly to the DNA backbone, trapping the replication fork. Because the *Ter* sites are arranged in two opposing sets of gates, they allow the two forks to enter the termination zone but prevent them from leaving it. Once the two forks are stopped at their meeting point, the remaining nucleotides are filled, the replisomes disassemble, and the two daughter chromosomes are completed.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Once the chromosome has been completely replicated, the two DNA copies move into two different cells during cell division. The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.
2. Helicase opens up the DNA-forming replication forks; these are extended in both directions.
3. Single-strand binding proteins coat the DNA around the replication fork to prevent rewinding of the DNA.
4. Topoisomerase binds at the region ahead of the replication fork to prevent supercoiling (over-winding).
5. Primase synthesizes RNA primers complementary to the DNA strand.
6. DNA polymerase III starts adding nucleotides to the 3'-OH (sugar) end of the primer.
7. Elongation of both the lagging and the leading strand continues.
8. RNA primers are removed and gaps are filled with DNA by DNA pol I.
9. The gaps between the DNA fragments are sealed by DNA ligase.

Key Terms

oriC

The primary origin of replication in *E. coli*.

Initiation

The stage of replication during which replication machinery assembles and DNA unwinding begins.

Elongation

The stage during which new DNA strands are synthesized.

Termination

The stage during which replication forks meet and DNA synthesis is completed.

Bidirectional Replication

Replication proceeding in two directions from an origin.

Ter Sites / Tus Protein

Directional trap system responsible for safely catching and halting converging replication forks at the termination zone of a

circular chromosome.

Check Your Understanding

1. What is *oriC*?
2. Why does bacterial replication begin at a specific origin?
3. Why is the presence of an A-T rich region at the *oriC* locus evolutionary advantageous for the initiation of replication?
4. What proteins participate in initiation?
5. Why are primers required during replication?
6. What happens during termination?
7. What would occur within an *E. coli* cell during division if the Ter sites were completely deleted from its chromosome? Would replication still conclude? Why or why not?

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Concepts of Biology. OpenStax CNX. May 18, 2016 <http://cnx.org/contents/s8Hh0oOc@9.10:2ousESf0@5/DNA-Replication>

1.2.4: DNA Replication in Eukaryotes

The essential steps of replication—initiation, elongation, and termination— are the same in eukaryotes as in bacteria. Both systems replicate DNA semi-conservatively, synthesize strands strictly in a 5' to 3' direction, require an RNA primer, and produce a continuous leading strand alongside a discontinuous lagging strand. However, because eukaryotic genomes are larger and linear, they have a more complex regulatory system and specialized enzymes.

Multiple Origins, Timing, and Licensing

The most obvious difference between bacterial and eukaryotic replication is genome scale. While *E. coli* replicates a single circular chromosome of ~4.6 million base pairs from a single origin, a human cell must copy 6 billion base pairs across 46 linear chromosomes during the S phase of the cell cycle. Moving at a standard eukaryotic fork velocity of 50 nucleotides per second, a single bacterial-style origin would take over a month to replicate one human chromosome. To compress this timeline into a 6-to-8-hour window, eukaryotic chromosomes contain thousands of origins firing simultaneously, creating replicons that eventually merge. To prevent gene-dosage imbalances, every origin must fire exactly once. Origins do not fire all at once; transcriptionally active, loosely packed euchromatin replicates early in the S phase, while tightly condensed, silent heterochromatin (like centromeres and telomeres) is delayed until late S phase.

To make sure that no origin fires twice during a single cell cycle, eukaryotes divide the process into two biochemical phases known as structural licensing and active firing. During the **structural licensing** phase, cyclin-dependent kinase (CDK) activity is kept intentionally low. This specific biochemical environment allows the six-protein **Origin Recognition Complex (ORC)** to bind directly to the replication origin, where it coordinates with specialized loading chaperones, Cdc6 and Cdt1, to deposit the inactive replicative helicase, the **MCM2-7 complex**, onto the DNA to complete the Pre-Replication Complex (pre-RC). Once the cell crosses the boundary into the active firing phase of S phase, CDK levels spike dramatically. This sudden surge activates key protein kinases that phosphorylate the pre-loaded MCM complex and recruit accessory proteins, assembling the active helicase motor. Simultaneously, this high CDK activity targets and destroys the loading chaperones while physically ejecting the ORC from the origin site. Because these essential loading factors are permanently disabled, no new helicases can be loaded onto an origin until the cell completely finishes division and returns to the next G₁ phase, effectively preventing over-replication.

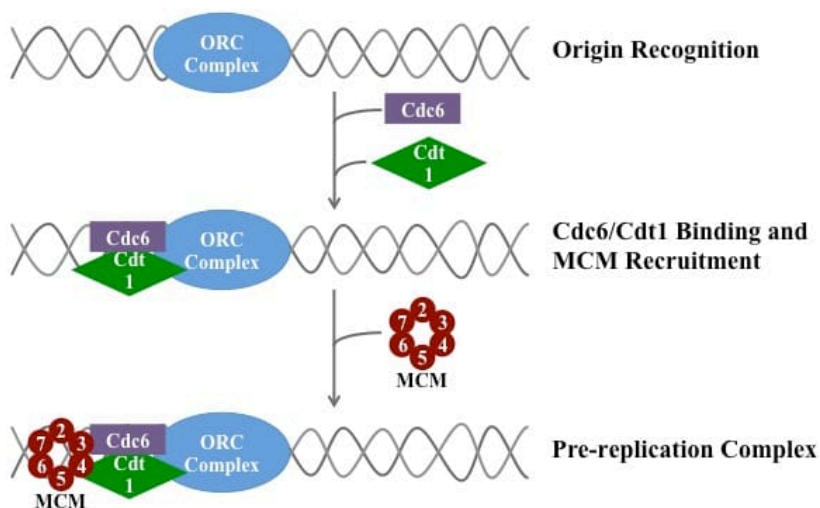


Figure 2.18 Loading of the pre-replication complex.

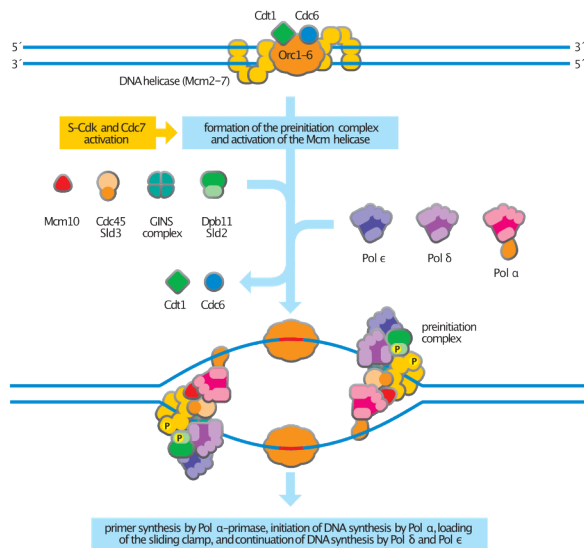


Figure 2.19 The activation of S-Cdks and Cdc7 triggers origin activation by promoting the formation of the pre-replication complex, a large group of proteins, including the DNA polymerases, that assembles at the origin, probably through an interaction with the Mcm2-7 complex. Some of the components of the preinitiation complex bind as dimers (Dpb1-Sld2 and Cdc45-Sld3) and tetramers (the GINS complex) and it is likely that these core proteins of the complex, together with Mcm10, bind simultaneously. Formation of the preinitiation complex results in the displacement of Cdc6 and Cdt1 from the ORC, which remains at the origin. Preinitiation complex formation activates the Mcm helicase, leading to unwinding of the DNA helix. On activation, Mcm subunits may be lost (not illustrated, for simplicity) or rearranged. RPA proteins (not shown) bind the single-stranded DNA, preventing its reannealing, and primase is recruited to begin primer synthesis and DNA replication. Synthesis of the leading strand begins first. The Mcm2-7 helicase and many components of the preinitiation complex travel out from the origin with the replication fork.

Replicating Through Chromatin and the Eukaryotic Replisome

Remember, eukaryotic DNA is wrapped around histone proteins to form nucleosomes. To advance, the eukaryotic replication fork must "remodel" the chromatin. Ahead of the fork, the histone chaperone **FACT** destabilizes the nucleosome, while the chaperone **Asf1** handles the displaced histone cores. Behind the advancing polymerases, the chaperone **CAF-1** captures these recycled parental histones, combines them with newly synthesized histones, and re-deposits them onto both daughter molecules.

While bacteria rely primarily on DNA Polymerase III to build both strands, the eukaryotic replisome uses three different polymerases for replication:

1. **DNA Polymerase α (alpha):** Operates in a complex with primase. Primase builds a ~10-nucleotide RNA primer, and Pol α extends it with ~20 DNA bases. Because Pol α lacks proofreading capabilities, it quickly hands the strand off to high-fidelity enzymes via a process called **polymerase switching**.
2. **DNA Polymerase ϵ (epsilon):** Is responsible for the continuous synthesis of the leading strand.
3. **DNA Polymerase δ (delta):** Is responsible for the discontinuous synthesis of the lagging strand.

The nucleosome barrier impacts the lagging strand replication. While bacterial Okazaki fragments are between 1,000 and 2,000 nucleotides, eukaryotic Okazaki fragments are much shorter, spanning just 100 to 200 nucleotides. Because DNA is wrapped around a histone core every 147 base pairs, the lagging strand polymerase can only synthesize the space between nucleosomes before running into a roadblock, necessitating thousands of successive priming events.

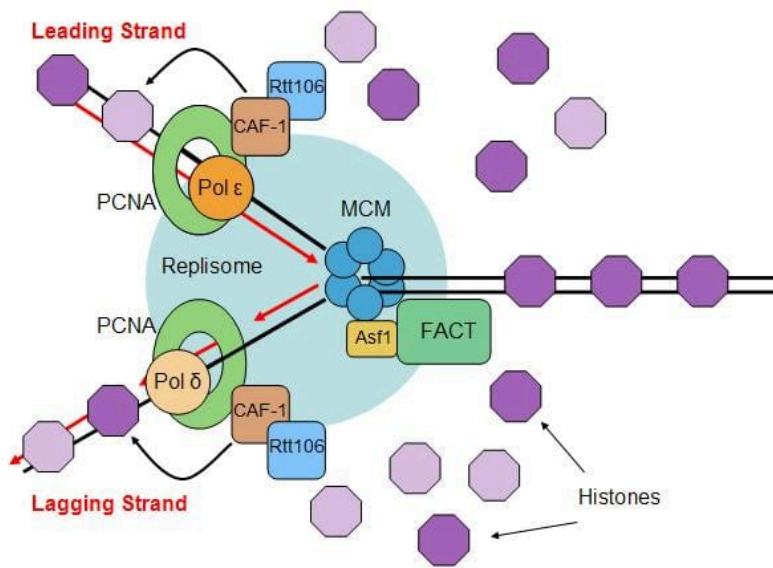


Figure 2.20 Depiction of replication through histones. Histones are removed from DNA by the FACT complex and Asf1. Histones are reassembled onto newly replicated DNA after the replication fork by CAF-1 and Rtt106.

Telomere Replication

The biggest difference between bacterial and eukaryotic replication comes from chromosome geometry. Because bacterial chromosomes are circular, replication forks loop continuously until they meet, leaving no exposed edges. Eukaryotic chromosomes are linear, which creates the **end-replication problem** on the lagging strand. Because DNA polymerases can only add nucleotides to a free 3'-OH group, the lagging strand relies on RNA primers to initiate each short Okazaki fragment. When the final RNA primer at the absolute outermost tip of the chromosome is degraded, there is no upstream DNA strand available for a polymerase to bind to and fill the gap, so the lagging strand template remains short and unpaired at its end. Left unchecked, chromosomes would progressively shorten with every single round of cell division, eventually eroding genetic information.

Eukaryotic cells solve this problem by capping their chromosome tips with non-coding segments called **telomeres**. Telomeres consist of simple, highly repetitive DNA sequences—in humans, the six-base sequence **5'-TTAGGG-3'** is repeated hundreds to thousands of times—acting as disposable sequences. The maintenance of the chromosome ends is done by a specialized *ribonucleoprotein* enzyme called **telomerase**. Telomerase contains an intrinsic RNA molecule that binds to the exposed 3' overhang of the parental DNA and serves as a template to extend the parental template strand. Once telomerase has lengthened the end, standard primase and DNA Polymerase δ can bind to the newly extended platform to synthesize a complementary lagging strand sequence.

In humans, telomerase is highly active in germ cells and embryonic stem cells, but is developmentally switched off in standard adult body cells (somatic cells). As a result, somatic telomeres shorten incrementally with every round of mitosis, associated with natural aging and cellular senescence. Conversely, while telomere shortening acts as a safety mechanism to prevent mutated cells from dividing indefinitely, roughly 90% of human cancer cells acquire mutations that pathologically reactivate telomerase. By restoring telomerase activity, cancer cells stabilize their chromosome tips, gaining a state of "immortality" that allows them to divide uncontrollably and metastasize.

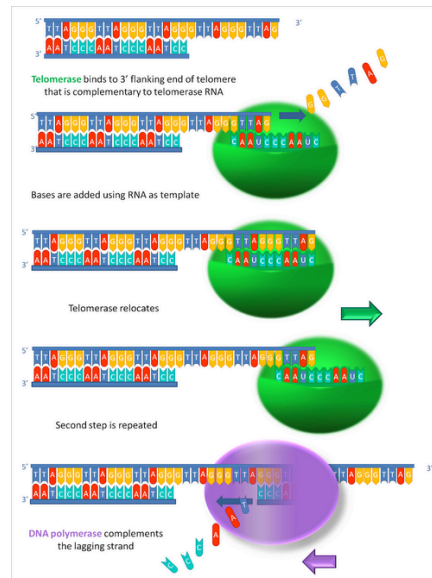


Figure 2.21 The ends of linear chromosomes are maintained by the action of the telomerase enzyme.



Figure 2.22 Elizabeth Blackburn, 2009 Nobel Laureate, is the scientist who discovered how telomerase works. (credit: US Embassy Sweden)

Key Terms

S Phase

The stage of the cell cycle during which DNA replication occurs.

Origin Licensing

The process that ensures origins are activated only once per cell cycle.

DNA Polymerase α

A polymerase involved in initiation of DNA synthesis.

DNA Polymerase δ

A major lagging-strand polymerase.

DNA Polymerase ϵ

A major leading-strand polymerase.

Polymerase Switching

The handoff mechanism where the error-prone, initiating DNA Polymerase dissociates from the template and is replaced by high-fidelity replicative polymerases.

FACT Complex

A specialized histone chaperone that physically disassembles nucleosomes ahead of the moving replication fork and instantly reassembles them behind it.

End-Replication Problem

The inability of conventional DNA polymerases to completely replicate the ends of linear chromosomes.

Telomere

A repetitive DNA sequence located at chromosome ends.

Telomerase

An enzyme that extends telomeres using an RNA template.

Check Your Understanding

1. Why do eukaryotic chromosomes require multiple origins of replication while bacteria function with just one?
2. During which phase of the cell cycle does replication occur?
3. What is origin licensing, and how does the cell prevent relicensing?
4. Why does chromatin create additional challenges for replication?
5. How are histones handled during replication?
6. Which polymerases primarily synthesize the leading and lagging strands in eukaryotes?
7. Why are eukaryotic Okazaki fragments shorter than bacterial fragments?
8. What problem arises because eukaryotic chromosomes are linear?
9. Why do chromosomes become shorter after replication?
10. What are telomeres?
11. How do telomeres protect chromosomes?
12. In which cell types is telomerase most active?
13. Why is telomerase important in cancer biology?

References

Unless otherwise noted, images on this page are licensed under CC-BY 4.0 by OpenStax.

OpenStax, Concepts of Biology. OpenStax CNX. May 18, 2016 <http://cnx.org/contents/s8Hh0oOc@9.10:2ousESf0@5/DNA-Replication>

Jaskelioff et al., 2011 Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice. Nature 469: 102-7.

1.3: TRANSCRIPTION

1.3.1: Transcription

1.3.2: Bacterial Transcription

1.3.3: Eukaryotic Transcription

1.3.4: RNA Processing

1.3.1: Transcription

In both prokaryotes and eukaryotes, the primary function of DNA is to provide the information needed to construct the proteins necessary so that the cell can perform all of its functions. To do this, a segment of the DNA, called a **gene**, is “read” or *transcribed* into a **messenger RNA molecule (mRNA)**. 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 3.1), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.

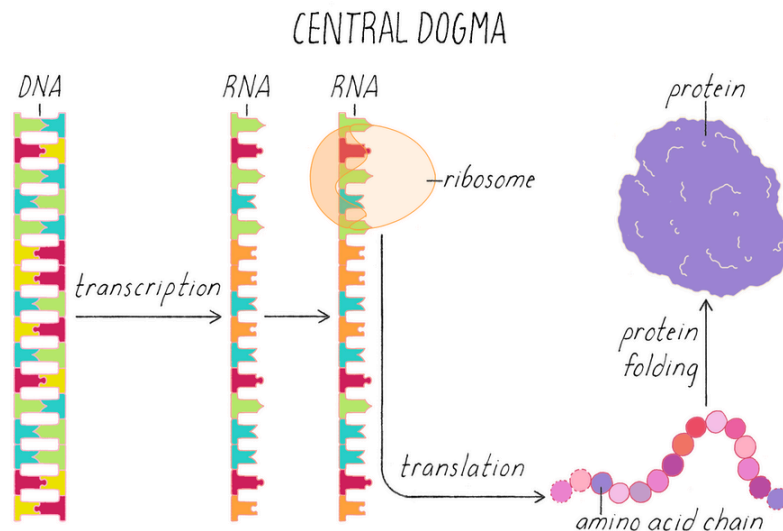


Figure 3.1 The central dogma states that DNA encodes RNA, which in turn encodes protein.

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.

Types of RNA

The process of transcription does not only produce messenger RNA. There are three major types of RNA that play a role in the central dogma. The first major type is messenger RNA (mRNA), which serves as the temporary, disposable copy of a gene. It is a single-stranded molecule that carries genetic information to the cellular machinery where proteins are assembled. Because mRNA is designed to be a temporary messenger, it has a relatively short lifespan; it is read a few times by the cell's machinery and then rapidly degraded and recycled.

To turn the code carried by mRNA into a physical protein, the cell relies on **transfer RNA (tRNA)** and **ribosomal RNA (rRNA)**. Shaped like a cloverleaf, tRNA acts as a translator. It is responsible for bringing individual amino acids to the protein assembly site. At one end, a tRNA molecule carries a specific amino acid, while its opposite end features a distinct three-nucleotide sequence that base-pairs with a complementary sequence on the mRNA strand, ensuring that the correct amino acid

is placed into the growing protein strand. The protein building "machine" is called the ribosome and is composed of proteins and **ribosomal RNA (rRNA)**. rRNA is a **ribozyme** or an RNA molecule that functions as an active enzyme. The rRNA within the ribosome actively catalyzes the formation of peptide bonds, which physically link amino acids together into a functional protein chain.

Along with the major types of RNA, cells produce many "non-coding" RNAs that do not carry instructions for proteins but instead perform essential structural and regulatory maintenance tasks. For example, small nuclear RNA (snRNA) is found in the eukaryotic nucleus, where it combines with proteins to form a complex called the spliceosome, which you will learn about later on in this unit. Additionally, ultra-short molecules known as microRNA (miRNA) and small interfering RNA (siRNA) act as fine-tuning knobs for the cell. By binding complementarily to specific mRNA transcripts floating in the cytoplasm, they can block protein production or target the mRNA for immediate destruction, effectively controlling gene expression levels.

Note: Why Uracil Instead of Thymine in RNA?

One of the key differences between DNA and RNA is that RNA uses uracil instead of thymine, and the reason for that comes down to the tradeoff between stability and energy efficiency. Remember, DNA needs to be stable over long periods of time, while mRNA is transient. Thymine is essentially a uracil with a methyl group added, making thymine more stable, but adding the methyl group is energetically costly for the cell. Because mRNA transcripts are temporary, short-lived messages meant to be read a few times and then immediately recycled, it would be incredibly wasteful for the cell to spend massive amounts of energy adding methyl tags to millions of temporary RNA bases.

Another reason for the difference has to do with DNA repair mechanisms. Cytosine (C) naturally decomposes over time into uracil (U) through a spontaneous chemical process called *deamination*. If DNA naturally used Uracil, repair enzymes in the cell wouldn't be able to tell the difference between a "correct" uracil or a "damaged" cytosine, leading to genetic mutations. By using thymine, the cell instantly knows that any uracil found in DNA is a mistake that needs to be fixed.

clipboard_e6014ff96c24c702428cfc7a8f5d1653d.png

Figure 2. The difference between thymine and uracil.

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**, similar to a replication 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 3.2).

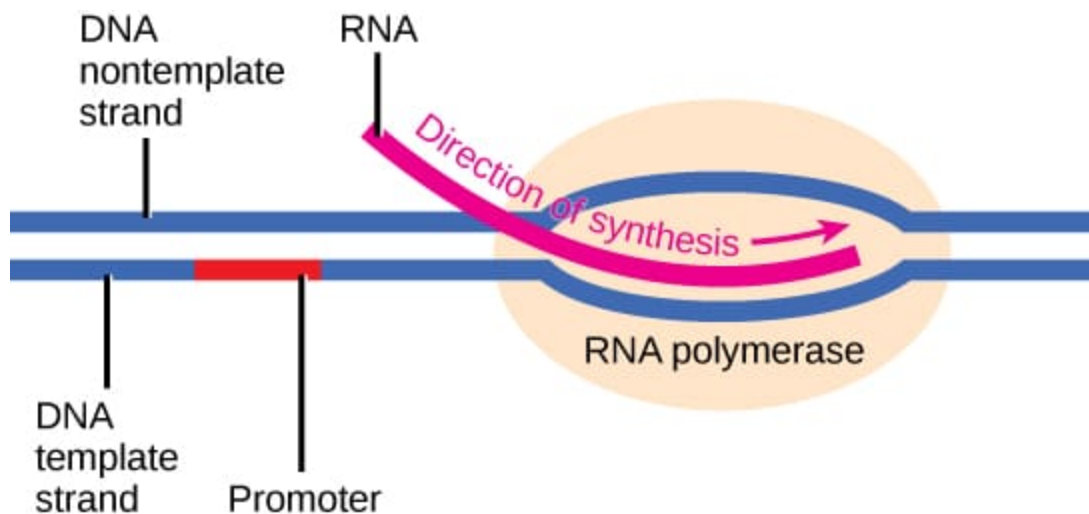


Figure 3.2 The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

Elongation

Transcription always proceeds from one of the two DNA strands, which is called the **template strand**, or **non-coding strand**. The mRNA product is *complementary* to this template. Consequently, the mRNA is almost identical to the other DNA strand, known interchangeably as the non-template or **coding strand**, with the sole exception that RNA contains uracil (U) in place of thymine (T). We call it the *coding* strand because its linear nucleotide sequence matches the actual genetic "code" of the final mRNA transcript used by the ribosome to build proteins.

The primary enzyme responsible for transcription is **RNA polymerase**. Unlike DNA polymerase, RNA polymerase can initiate the building of a new strand from scratch, it doesn't require a pre-existing nucleotide chain. Once bound to the promoter, RNA polymerase unzips the DNA to form the transcription bubble, then moves down the strand reading in a 3' to 5' direction, matching free RNA nucleotides to the exposed DNA template strand. It "reads" the template strand 3' to 5', and "writes" the mRNA 5' to 3'. As the enzyme passes, the DNA double helix zips back together behind it, allowing a single-stranded RNA molecule to cleanly peel away from the complex.

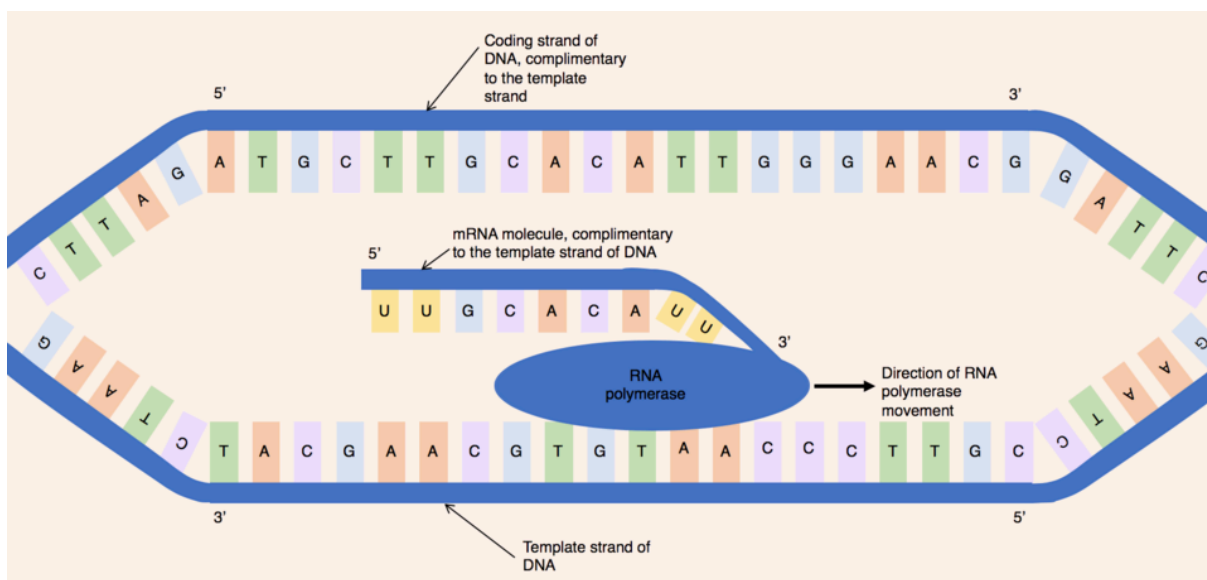


Figure 3.3 During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.

Termination

Transcription termination marks the official end of the transcription process, signaling the newly built RNA molecule to detach and the enzyme machinery to release the DNA template. Once RNA polymerase encounters a specific, encoded sequence at the end of a gene called a **terminator**, it triggers a series of conformational changes that destabilize the entire complex. This slowdown breaks the hydrogen bonds holding the hybrid DNA-RNA strand together, causing the freshly synthesized single-stranded RNA transcript to cleanly peel away. Simultaneously, the RNA polymerase enzyme releases its grip on the parental template, allowing the opened DNA bubble to zip back together into its stable double-helix structure.

Gene Orientation

A common point of confusion is visualizing how genes are physically arranged and oriented along a double-stranded chromosome. First, a gene is inherently single-stranded; while genomic DNA is a double helix, a single gene resides on only one of those two strands. The opposing strand serves purely as the complementary partner for that specific gene and does not encode the same biological instruction. Second, either the "top" or the "bottom" strand can serve as the template. Cells do not exclusively use one continuous strand of a chromosome to transcribe all of their genes. Because promoters have distinct physical directions, some genes are read left-to-right using the bottom strand as a template, while neighboring genes might be read right-to-left using the top strand as a template.

The choice of which strand acts as the template is determined completely by the orientation of the gene's promoter sequence. Because RNA polymerase must always bind a promoter and read the designated template strand in a 3' to 5' direction, the "top" strand might serve as the template for one gene, while the "bottom" strand acts as the template for an adjacent gene. Ultimately, transcription is a localized process, and the two strands of a double helix never simultaneously encode the same active genetic message.

When looking at a static diagram in a textbook, it is easy to assume that transcription must always move from left to right across the page. **This is false.** Because RNA polymerase cares only about chemical polarity, not the layout of a textbook page, it can transcribe in either physical direction. If a gene's promoter points to the left, RNA polymerase will happily latch onto the opposite strand and transcribe from right to left!

Key Terms

Gene

The specific DNA sequence that contains the instructions for building a functional biological molecule—most commonly a protein or an RNA molecule.

Transcription

The synthesis of RNA using a DNA template.

Messenger RNA (mRNA)

RNA that carries genetic information from DNA to ribosomes.

Ribosomal RNA (rRNA)

RNA that forms the structural and catalytic core of ribosomes.

Transfer RNA (tRNA)

RNA that delivers amino acids during protein synthesis.

Ribozyme

An RNA molecule capable of catalyzing chemical reactions.

Promoter

The specific DNA sequence upstream (before) of a gene that acts as the transcription "start signal".

Transcription Bubble

The region of unwound DNA where transcription occurs.

RNA Polymerase

The enzyme responsible for RNA synthesis.

Template Strand

The DNA strand used to direct RNA synthesis.

Coding Strand

The DNA strand whose sequence resembles the RNA transcript.

Check Your Understanding

1. What is the primary function of mRNA?
2. What is transcription?
3. Which enzyme performs transcription?
4. What is the difference between the template strand and the coding strand?
5. Why does RNA contain uracil rather than thymine?
6. In which direction is RNA synthesized?
7. What is a transcription bubble?
8. Why does RNA polymerase not require a primer?
9. Do all genes encode proteins? Explain.
10. What types of RNA can be produced by transcription?
11. RNA is considered one of the most versatile molecules in biology, why do you think that is?

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [1.3.1: Transcription](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [9.3: Transcription](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

1.3.2: Bacterial Transcription

Initiation of Transcription in Bacteria

Bacteria do not have membrane-enclosed nuclei. Therefore, the processes of transcription, translation, and mRNA degradation can all occur simultaneously. The intracellular level of a bacterial protein can quickly be amplified by multiple transcription and translation events that occur concurrently on the same DNA template. Bacterial genomes are very compact. To maximize efficiency, genes with related functions are often grouped together on the DNA under the control of a single promoter. This functional cluster is called an **operon**. Because the entire operon is transcribed together in a single run, the resulting mRNA transcript covers more than one gene or cistron (a coding sequence for a single protein). These **polycistronic mRNAs** are then translated to produce multiple distinct types of proteins at the same time.



Figure 3.4 Structure of the genes in the alpha operon of *e. coli*. All of these genes are under the control of one promoter and would be transcribed into one mRNA, and then translated into separate proteins.

Our discussion here will exemplify transcription by describing this process in *Escherichia coli*, a well-studied model species. An understanding of *E. coli* transcription can be applied to virtually all bacterial species.

Bacterial RNA Polymerase

Bacteria use the same RNA polymerase to transcribe all of their genes. In *E. coli*, the polymerase is composed of six polypeptide subunits: two identical α subunits, one β subunit, one β' subunit, and a small stabilizing ω subunit. These subunits make up the core enzyme and assemble every time a gene is transcribed, and they disassemble once transcription is complete. Each subunit has a unique role; the two α -subunits are necessary to assemble the polymerase on the DNA; the β -subunit binds to the ribonucleoside triphosphate that will become part of the nascent mRNA molecule; the β' subunit binds the DNA template strand; and the ω subunit stabilizes the core enzyme. The sixth subunit, σ (**sigma**) **factor**, is involved only in transcription initiation. It confers transcriptional specificity so that the polymerase begins to synthesize mRNA at the correct initiation site. Without the σ subunit, the core enzyme would transcribe from random sites and would produce mRNA molecules that specified protein gibberish. The polymerase comprised of all five subunits is called the **holoenzyme**.

Because the core enzyme handles the actual synthesis of all RNA molecules, bacteria require only a single type of RNA polymerase to transcribe all of their genes. To change which genes are being transcribed, the cell simply switches out the σ factor. The primary σ factor (such as σ^{70} in *E. coli*) directs the holoenzyme to standard housekeeping genes needed for everyday growth. However, under environmental stress, such as starvation or extreme heat, the cell can synthesize alternative σ factors. These alternative factors displace the primary σ factor, shifting the holoenzyme's focus to a completely new set of promoters to help the cell survive.

Bacterial Promoters

A promoter is a DNA sequence onto which the transcription machinery, including RNA polymerase, binds and initiates transcription. 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 the time, some of the time, or infrequently. Although promoters vary among bacterial genomes, a few elements are evolutionarily conserved in many species. At the -10 and -35 regions upstream of the initiation site, there are two **promoter consensus sequences**, or regions that are similar across all promoters and across various bacterial species (Figure 3.5). The -10 sequence, called the **-10 region**, or **Pribnow Box**, has the consensus sequence TATAAT. The **-35 sequence** has the consensus sequence TTGACA. These consensus sequences are recognized and bound by σ . Once this interaction is made, the subunits of the core enzyme bind to the site. The A-T-rich -10 region facilitates unwinding of the DNA template. Before the holoenzyme can move down the DNA, it often goes through a trial-and-error process called **abortive initiation**. Because the holoenzyme is tightly bound to the promoter, it repeatedly synthesizes and releases short RNA fragments (approximately 10 nucleotides long) without moving forward. The initiation phase officially ends only when the enzyme successfully builds a slightly longer transcript, breaks its grip on the promoter and releases the sigma

factor.

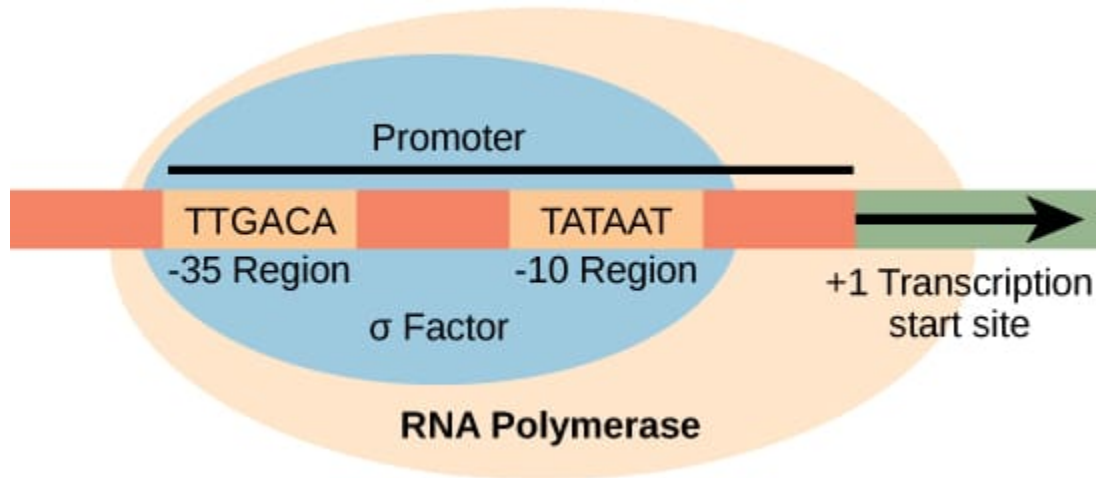


Figure 3.5 The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start site. The σ subunit dissociates from the polymerase after transcription has been initiated.

Link to Learning

Link to Learning

View this [MolecularMovies animation](#) to see the transcription process as it happens in the cell.

Elongation

The transcription elongation phase begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core enzyme to proceed along the DNA template, synthesizing mRNA in the 5' to 3' direction at a rate of approximately 40 nucleotides per second. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. The hydrogen bonds between the template DNA and the newly formed RNA strand are too weak to hold the heavy transcription machinery together on their own. Instead, the RNA polymerase itself acts as a bridge, tightly gripping both the DNA template and the nascent RNA. By physically anchoring these strands within its protein core, the enzyme prevents the fragile complex from falling apart prematurely during elongation.

Bacterial Termination Signals

Once a gene is transcribed, the RNA 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. One is protein-based and the other is RNA-based. **Rho-dependent termination** is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination, also called intrinsic termination, is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a stable hairpin that causes the polymerase to stall as soon as it begins to transcribe a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

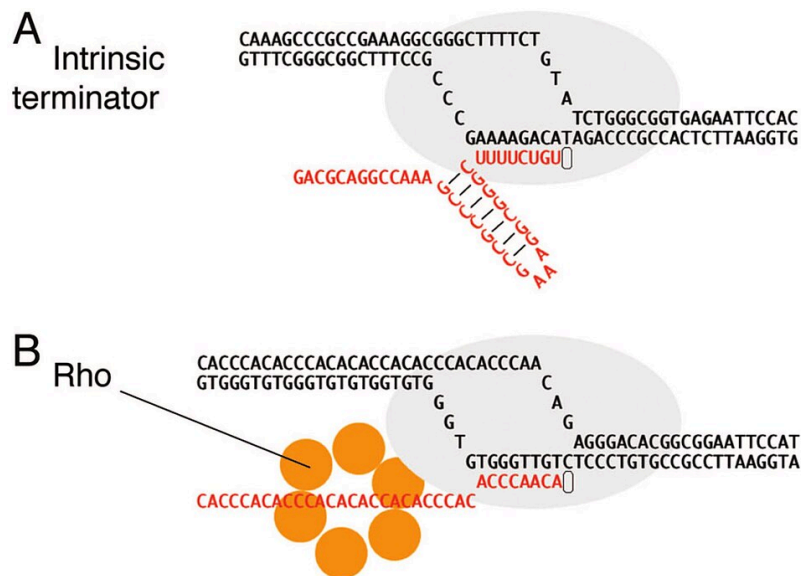


Figure 3.6 Comparison of Rho-independent (A) and Rho-dependent (B) termination

Upon termination, the process of transcription is complete. By the time termination occurs, the bacterial transcript would already have been used to begin synthesis of numerous copies of the encoded protein because transcription and translation can occur concurrently. The unification of transcription, translation, and even mRNA degradation is possible because all of these processes occur in the same 5' to 3' direction, and because there is no membranous compartmentalization in the prokaryotic cell (Figure 15.9). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

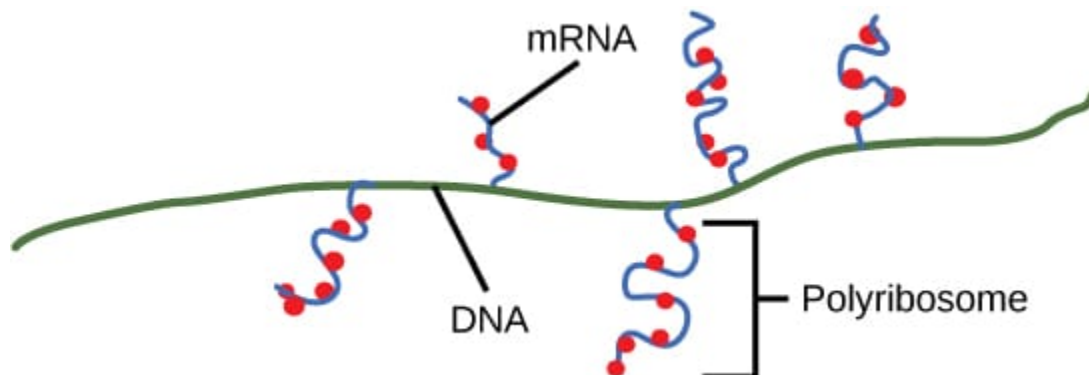


Figure 3.7 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Key Terms

Polycistronic mRNA

An mRNA molecule that encodes multiple proteins.

Operon

A cluster of genes transcribed together from a single promoter.

Sigma (σ) Factor

A protein that directs bacterial RNA polymerase to promoter sequences.

RNA Polymerase Holoenzyme

The bacterial RNA polymerase core enzyme bound to a sigma factor.

Promoter

A DNA sequence where transcription begins.

-10 Region (Pribnow Box)

An AT-rich promoter sequence recognized during transcription initiation.

-35 Region

A conserved promoter sequence recognized by sigma factors.

Rho-Independent Termination

Termination caused by the formation of an RNA hairpin followed by a string of uracils.

Rho-Dependent Termination

Termination requiring the Rho helicase protein.

Check Your Understanding

1. Why do bacteria require only one RNA polymerase?
2. What is the function of a sigma factor?
3. What are the -10 and -35 promoter regions?
4. What happens to the sigma factor after transcription begins?
5. Compare rho-independent and rho-dependent termination.
6. Why can transcription and translation occur simultaneously in bacteria?
7. What is a polycistronic mRNA?
8. Why are operons beneficial to bacterial cells?

1.3.3: Eukaryotic Transcription

Prokaryotes and eukaryotes perform fundamentally the same process of transcription, with a few key differences. The most important difference is due to the membrane bound nucleus of eukaryotes. With the genes located in the nucleus, the eukaryotic cell must be able to transport its mRNA to the cytoplasm and must protect it from degrading before it is translated. Eukaryotes also use three different polymerases that each transcribe a different subset of genes. Also, unlike prokaryotes, eukaryotic mRNAs are usually monogenic, meaning they specify just one single gene.

Initiation of Transcription in Eukaryotes

Unlike the bacterial polymerase that can bind to a DNA template on its own, eukaryotes require several other proteins, called **transcription factors (TFs)**, to first bind to the promoter region and then to help recruit the appropriate polymerase.

The Three Eukaryotic RNA Polymerases

The features of eukaryotic mRNA synthesis are markedly more complex than those of bacteria. Instead of a single polymerase comprising six subunits, the eukaryotes have three polymerases that are each made up of 10 subunits or more. Each eukaryotic polymerase also requires a distinct set of transcription factors to bring it to the DNA template.

RNA polymerase I is located in the *nucleolus*, a specialized nuclear substructure in which ribosomal RNA (rRNA) is transcribed, processed, and assembled into ribosomes (Table 15.1). The rRNA molecules are considered *structural RNAs* because they have a cellular role but are not translated into protein. The rRNAs are components of the ribosome and are essential to the process of translation. RNA polymerase I synthesizes all of the rRNAs from the tandemly duplicated set of 18S, 5.8S, and 28S ribosomal genes. (Note that the “S” designation applies to “Svedberg” units, a nonadditive value that characterizes the speed at which a particle sediments during centrifugation.)

RNA polymerase II is located in the nucleus and synthesizes all protein-coding nuclear pre-mRNAs. Eukaryotic pre-mRNAs undergo extensive processing after transcription but before translation. For clarity, this module’s discussion of transcription and translation in eukaryotes will use the term “mRNAs” to describe only the mature, processed molecules that are ready to be translated. RNA polymerase II is responsible for transcribing the overwhelming majority of eukaryotic genes.

RNA polymerase III is also located in the nucleus. This polymerase transcribes a variety of structural RNAs that includes the 5S pre-rRNA, transfer pre-RNAs (pre-tRNAs), and small nuclear pre-RNAs. The tRNAs have a critical role in translation; they serve as the “adaptor molecules” between the mRNA template and the growing polypeptide chain. Small nuclear RNAs have a variety of functions, including “splicing” pre-mRNAs and regulating transcription factors.

A scientist characterizing a new gene can determine which polymerase transcribes it by testing whether the gene is expressed in the presence of α -amanitin, an oligopeptide toxin produced by the fly agaric toadstool mushroom and other species of *Amanita*. Interestingly, the α -amanitin affects the three polymerases very differently (Table 3.1). RNA polymerase I is completely insensitive to α -amanitin, meaning that the polymerase can transcribe DNA in vitro in the presence of this poison. RNA polymerase III is moderately sensitive to the toxin. In contrast, RNA polymerase II is extremely sensitive to α -amanitin. The toxin prevents the enzyme from progressing down the DNA, and thus inhibits transcription. Knowing the transcribing polymerase can provide clues as to the general function of the gene being studied. Because RNA polymerase II transcribes the vast majority of genes, we will focus on this polymerase in our subsequent discussions about eukaryotic transcription factors and promoters.

Table 3.1 Locations, Products, and Sensitivities of the Three Eukaryotic RNA Polymerases

RNA Polymerase	Cellular Compartment	Product of Transcription	α -Amanitin Sensitivity
I	Nucleolus	All rRNAs except 5S rRNA	Insensitive
II	Nucleus	All protein-coding nuclear pre-mRNAs	Extremely sensitive
III	Nucleus	5S rRNA, tRNAs, and small nuclear RNAs	Moderately sensitive

RNA Polymerase II Promoters and Transcription Factors

Eukaryotic promoters are much larger and more intricate than bacterial promoters. However, both have a sequence similar to the -10 sequence of bacteria. In eukaryotes, this sequence is called the **TATA box**, and has the consensus sequence TATAAA on the coding strand. It is located at -25 to -35 bases relative to the initiation (+1) site (Figure 3.8). This sequence is not identical to the *E. coli* -10 box, but it conserves the A-T rich element. The thermostability of A-T bonds is low and this helps the DNA template to locally unwind in preparation for transcription.

Instead of the simple σ factor that helps bind the prokaryotic RNA polymerase to its promoter, eukaryotes assemble a complex of transcription factors required to recruit RNA polymerase II to a protein coding gene because eukaryotic RNA polymerases cannot recognize the promoter on their own. Transcription factors that bind to the promoter are called **basal transcription factors**. These basal factors are all called TFII (for Transcription Factor/polymerase II) plus an additional letter (A-J). The core complex is **TFIID**, which includes a TATA-binding protein (TBP). The other transcription factors systematically fall into place on the DNA template, with each one further stabilizing the pre-initiation complex and contributing to the recruitment of RNA polymerase II.

Visual Connection

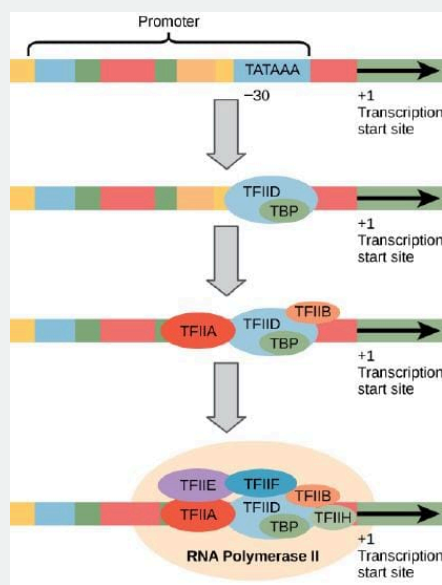


Figure 3.8 A generalized promoter of a gene transcribed by RNA polymerase II is shown. Transcription factors recognize the promoter. RNA polymerase II then binds and forms the transcription initiation complex.

Some eukaryotic promoters also have a conserved **CAAT box** (GGCCAATCT) at approximately -80. Further upstream of the TATA box, eukaryotic promoters may also contain one or more **GC-rich boxes** (GGCG) or octamer boxes (ATTTGCAT). These elements bind cellular factors that increase the efficiency of transcription initiation and are often identified in more “active” genes that are constantly being expressed by the cell.

Basal transcription factors are crucial in the formation of a **pre-initiation complex** on the DNA template that subsequently recruits RNA polymerase II for transcription initiation. The complexity of eukaryotic transcription does not end with the polymerases and promoters. An army of other transcription factors, which bind to upstream enhancers and silencers, also help to regulate the frequency with which pre-mRNA is synthesized from a gene. Enhancers and silencers affect the efficiency of transcription but are not necessary for transcription to proceed.

Promoter Structures for RNA Polymerases I and III

The processes of bringing RNA polymerases I and III to the DNA template involve slightly less complex collections of transcription factors, but the general theme is the same. The conserved promoter elements for genes transcribed by polymerases I and III differ from those transcribed by RNA polymerase II. RNA polymerase I transcribes genes that have two GC-rich promoter sequences in the -45 to +20 region. These sequences alone are sufficient for transcription initiation to occur, but promoters with additional sequences in the region from -180 to -105 upstream of the initiation site will further enhance initiation. Genes that are

transcribed by RNA polymerase III have upstream promoters or promoters that occur within the genes themselves.

Eukaryotic transcription is a tightly regulated process that requires a variety of proteins to interact with each other and with the DNA strand. Although the process of transcription in eukaryotes involves a greater metabolic investment than in prokaryotes, it ensures that the cell transcribes precisely the pre-mRNAs that it needs for protein synthesis.

Evolution Connection

The Evolution of Promoters

The evolution of genes may be a familiar concept. Mutations can occur in genes during DNA replication, and the result may or may not be beneficial to the cell. By altering an enzyme, structural protein, or some other factor, the process of mutation can transform functions or physical features. However, eukaryotic promoters and other gene regulatory sequences may evolve as well. For instance, consider a gene that, over many generations, becomes more valuable to the cell. Maybe the gene encodes a structural protein that the cell needs to synthesize in abundance for a certain function. If this is the case, it would be beneficial to the cell for that gene's promoter to recruit transcription factors more efficiently and increase gene expression.

Scientists examining the evolution of promoter sequences have reported varying results. In part, this is because it is difficult to infer exactly where a eukaryotic promoter begins and ends. Some promoters occur within genes; others are located very far upstream, or even downstream, of the genes they are regulating. However, when researchers limited their examination to human core promoter sequences that were defined experimentally as sequences that bind the pre-initiation complex, they found that promoters evolve even faster than protein-coding genes.

It is still unclear how promoter evolution might correspond to the evolution of humans or other complex organisms. However, the evolution of a promoter to effectively make more or less of a given gene product is an intriguing alternative to the evolution of the genes themselves.¹

Eukaryotic Elongation and Termination

Following the formation of the pre-initiation complex, the polymerase is released from the other transcription factors, and elongation is allowed to proceed as it does in bacteria with the polymerase synthesizing pre-mRNA in the 5' to 3' direction. As discussed previously, RNA polymerase II transcribes the major share of eukaryotic genes, so in this section we will focus on how this polymerase accomplishes elongation and termination.

Although the enzymatic process of elongation is essentially the same in eukaryotes and prokaryotes, the DNA template is considerably more complex. When eukaryotic cells are not dividing, their genes exist as chromatin. For mRNA synthesis to occur, the transcription machinery needs to move histones out of the way every time it encounters a nucleosome. This is accomplished by a special protein complex called FACT, which stands for "*facilitates chromatin transcription*." This complex pulls histones away from the DNA template as the polymerase moves along it. Once the pre-mRNA is synthesized, the FACT complex replaces the histones to recreate the nucleosomes.

The termination of transcription is different for the different polymerases. Unlike in bacteria, elongation by RNA polymerase II in eukaryotes takes place 1,000 to 2,000 nucleotides beyond the end of the gene being transcribed. This pre-mRNA tail is subsequently removed by cleavage during mRNA processing. On the other hand, RNA polymerases I and III require termination signals. Genes transcribed by RNA polymerase I contain a specific 18-nucleotide sequence that is recognized by a termination protein. The process of termination in RNA polymerase III involves an mRNA hairpin similar to rho-independent termination of transcription in prokaryotes.

Replication in Archaea

While organisms in the domain Archaea structurally resemble bacteria due to their lack of a membrane-bound nucleus, transcription is fundamentally eukaryotic in nature. In bacteria, a single relatively straightforward RNA polymerase core enzyme. In contrast, archaea use a single, complex RNA polymerase composed of 11 to 13 distinct protein subunits. Structurally and chemically, this multi-subunit enzyme is a homolog of eukaryotic RNA polymerase II. The amino acid sequences, three-dimensional spatial configurations, and catalytic mechanisms of individual archaeal subunits precisely map onto the

corresponding eukaryotic counterpart. Rather than using the classic bacterial -10 and -35 promoter sequences, archaeal promoters are structurally similar to those found in eukaryotes. Many archaeal promoters rely on three consensus sequences positioned upstream of the transcription start site. The first of these is the B Recognition Element (BRE), which is a conserved, purine-rich sequence located immediately upstream of the TATA box that assists in establishing the proper orientation of the transcription machinery. Directly adjacent to the BRE is the TATA box itself, located approximately 25 to 30 base pairs upstream of the transcription start site that is structurally identical to the eukaryotic TATA box. To identify the promoter and successfully initiate transcription, archaea do not use bacterial like sigma factors. Instead, they depend on a pathway that pre-dates but mirrors the eukaryotic pre-initiation complex. This system relies on two universal general transcription factors and the process begins with the TATA-Binding Protein (TBP). Next, Transcription Factor B (TFB) binds to the upstream BRE promoter sequence and the pre-bound TBP molecule. Once locked into place, TFB recruits and stabilizes the multi-subunit archaeal RNA polymerase, correctly positioning it over the transcription start site. Because of this specific role, TFB serves as the direct counterpart to the eukaryotic general transcription factor TFIIB. Together, TBP, TFB, and the multi-subunit RNA polymerase assemble to form a minimal, fully operational transcription initiation complex. Archaeal transcription is one line of evidence pointing to eukaryotic life originating from an archaeal ancestor.

Key Terms

RNA Polymerase I

Synthesizes most ribosomal RNA.

RNA Polymerase II

Synthesizes messenger RNA and many regulatory RNAs.

RNA Polymerase III

Synthesizes transfer RNA and several small RNAs.

General Transcription Factors

Proteins that recruit and position RNA Polymerase II at promoters.

TATA Box

A common promoter element recognized during transcription initiation.

Pre-Initiation Complex (PIC)

The assembly of RNA Polymerase II and general transcription factors at a promoter before transcription begins.

Check Your Understanding

1. Why is transcription more complex in eukaryotes than in bacteria?
2. What are the roles of RNA Polymerases I, II, and III?
3. What is the function of the TATA box?
4. Why are general transcription factors necessary?
5. What is the pre-initiation complex?
6. Why is chromatin remodeling required before transcription?
7. Why are transcription and translation separated in eukaryotic cells?

Footnotes

- ¹H Liang et al., "Fast evolution of core promoters in primate genomes," *Molecular Biology and Evolution* 25 (2008): 1239–44.

1.3.4: RNA Processing

After transcription, eukaryotic pre-mRNAs must undergo several processing steps before they can be translated. Eukaryotic (and prokaryotic) tRNAs and rRNAs also must be processed before they can function as components in the protein-synthesis machinery.

mRNA Processing

The eukaryotic pre-mRNA undergoes extensive processing before it is ready to be translated. Eukaryotic protein-coding sequences are not continuous, as they are in prokaryotes. The coding sequences, called **exons**, are interrupted by noncoding **introns**, which must be removed to make a *translatable* mRNA. The additional steps involved in eukaryotic mRNA maturation also create a molecule with a much longer half-life than a prokaryotic mRNA. Eukaryotic mRNAs last for several hours, whereas the typical *E. coli* mRNA lasts a few minutes.

Pre-mRNAs are first coated in RNA-stabilizing proteins; these protect the pre-mRNA from degradation while it is processed and exported out of the nucleus. The three most important steps of pre-mRNA processing are the addition of stabilizing and signaling factors at the 5' and 3' ends of the molecule, and the removal of the introns (Figure 3.9). In some cases, the mRNA transcript can be “edited” after it is transcribed.

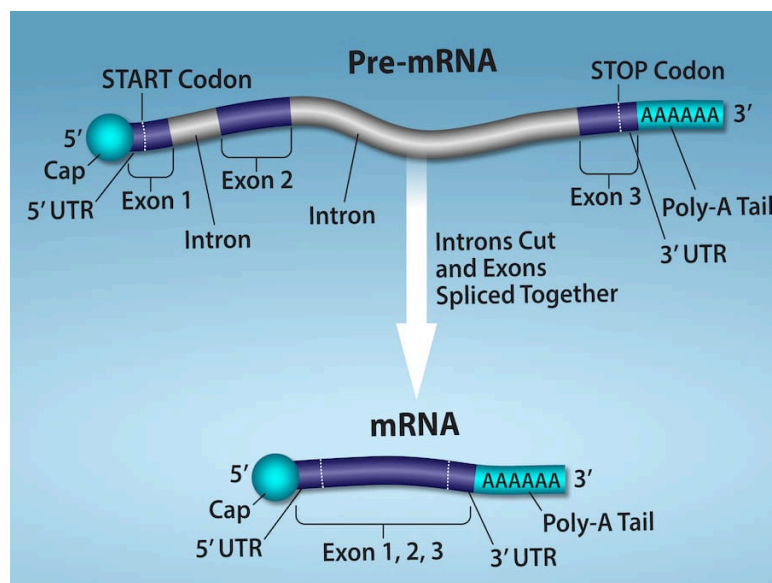


Figure 3.9 Eukaryotic pre-mRNA processing. Introns must be removed and exons joined to generate a functional mRNA. Nucleotides upstream of the translation START codon are part of the 5' untranslated region (5' UTR). Nucleotides downstream of the STOP codon form the 3' UTR. Both 5' and 3' UTRs are important for regulating translation initiation and mRNA stability. Credit: Rao, A., Ryan, K. Fletcher, S. and Tag, A. Department of Biology, Texas A&M University.

Evolution Connection

Real-Time Adaptation via Rewriting RNA in Cephalopods

The classic pathway of the Central Dogma dictates that genetic information flows predictably from DNA to RNA to protein, meaning that a gene's sequence directly determines the final protein structure. However, shell-less cephalopods, including octopuses, squid, and cuttlefish, completely bypass this rule through an extraordinary molecular process known as RNA editing. While mammals only edit a tiny handful of their mRNA bases, cephalopods actively rewrite tens of thousands of individual bases within their transcripts, particularly inside their complex nervous systems.

Instead of altering the sequence length, cephalopods have specialized enzymes called **ADARs** (Adenosine Deaminases Acting on RNA) that remove an amine group from an adenosine base, converting it into a molecule called inosine (I). Because the cell's translation machinery reads an Inosine as if it were a guanosine (G), this targeted deamination alters the triplet codon. As a result, a single unaltered gene can yield entirely different protein variations on the fly.

Recent marine genomics research reveals that cephalopods use this editing system as an emergency acclimation toolkit to survive sudden environmental threats like extreme temperature drops. When an octopus swims into deep, freezing ocean water, its cells rapidly alter more than 13,000 distinct mRNA sites. This real-time recoding modifies the physical shapes of critical neurological proteins, such as kinesin motor proteins and voltage-gated potassium channels, preventing the animal's nervous system from slowing down or freezing up in the deep cold.

This real-time physiological flexibility comes with a trade-off, however. In order for ADAR enzymes to recognize and edit a transcript, the surrounding pre-mRNA must fold into specific, rigid, double-stranded structures. To maintain these exact structural targets, cephalopods cannot tolerate natural, random mutations within the flanking DNA sequences. Consequently, they have effectively locked large portions of their genomic DNA evolution in place, exchanging long term evolution for the immediate benefits of mRNA editing.



5' Capping

While the pre-mRNA is still being synthesized, a modified guanine nucleotide called a **7-methylguanosine (m^7G) cap** is added to the 5' end of the growing transcript through a unique 5'-to-5' triphosphate linkage. This chemical modification acts as a shield, preventing cellular exonucleases from binding and prematurely degrading the newly made RNA strand. Beyond stability, eukaryotic translation initiation factors recognize and bind directly to this cap, allowing the small ribosomal subunit to correctly identify the front of the mRNA molecule, dock onto it, and begin scanning for the start codon.

3' Poly-A Tail

Once the elongation phase is complete, a highly specific cleavage complex identifies an **AAUAAA polyadenylation signal** near the end of the transcript. An endonuclease slices the pre-mRNA just downstream of this signal, shedding a GU-rich trailer sequence and generating a new 3' terminus. Immediately after this cleavage event, an enzyme called **poly-A polymerase** appends a long, continuous sequence of approximately 200 adenine nucleotides to the freshly cut end. This **poly-A tail** performs a dual protective and structural role. It prevents cellular enzymes from destroying the back end of the transcript, while simultaneously serving as a docking site for poly-A-binding proteins, which physically guide the fully mature mRNA through the nuclear pore complex and export it out into the cytoplasm for translation.

Once elongation is complete, the pre-mRNA is cleaved by an endonuclease between an AAUAAA consensus sequence and a GU-rich sequence, leaving the AAUAAA sequence on the pre-mRNA. Immediately after the cleavage event, an enzyme called **poly-A polymerase** adds a string of approximately 200 Adenine residues, called the **poly-A tail**. The poly-A tail prevents enzymes from

destroying the back end of the transcript and is also the binding site for a protein necessary for exporting the processed mRNA through nuclear pores into the cytoplasm for translation.

Pre-mRNA Splicing

Eukaryotic genes are composed of exons, which correspond to protein-coding sequences (*ex-on* signifies that they are expressed), and *intervening* sequences called introns (*int-ron* denotes their *intervening* role), which may be involved in gene regulation but are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins.

The discovery of introns came as a surprise to researchers in the 1970s who expected that pre-mRNAs would specify protein sequences without further processing, as they had observed in prokaryotes. The genes of higher eukaryotes very often contain one or more introns. These regions may correspond to regulatory sequences; however, the biological significance of having many introns or having very long introns in a gene is unclear. It is possible that introns slow down gene expression because it takes longer to transcribe pre-mRNAs with lots of introns. Alternatively, introns may be nonfunctional sequence remnants left over from the fusion of ancient genes throughout the course of evolution. This is supported by the fact that separate exons often encode separate protein subunits or domains. For the most part, the sequences of introns can be mutated without ultimately affecting the protein product.

All of a pre-mRNA's introns must be completely and precisely removed before protein synthesis. If the process errs by even a single nucleotide, the reading frame of the rejoined exons would shift, and the resulting protein would be dysfunctional. The process of removing introns and reconnecting exons is called **splicing** (Figure 3.10). Introns are removed and degraded while the pre-mRNA is still in the nucleus. Although the intron itself is noncoding, the beginning and end of each intron is marked with specific nucleotides: GU at the 5' end and AG at the 3' end of the intron. The splicing of pre-mRNAs is conducted by a complex of proteins and small nuclear RNAs (snRNAs) called the **spliceosome**. Crucially, the spliceosome functions as a **ribozyme**; the snRNA molecules, rather than the protein components, are responsible for recognizing the splice site sequences and catalyzing the chemical reactions.

The physical removal of an intron occurs through two successive transesterification reactions, which are chemical rearrangements that exchange one phosphodiester bond for another without consuming ATP. During the first transesterification, the spliceosome brings a specific adenine nucleotide within the center of the intron—known as the *branch point*—into close physical proximity with the 5' splice site. The 2'-hydroxyl (-OH) group of this branch-point adenine launches a nucleophilic chemical attack on the phosphodiester bond at the 5' GU boundary. This reaction severs the intron from Exon 1, causing the newly freed 5' end of the intron to loop backward and form a covalent bond with the branch-point adenine, rolling the intron into a lasso-like structure called a **lariat**. Immediately following this event, the second transesterification begins as the freshly exposed 3'-hydroxyl group of Exon 1 launches a subsequent chemical attack on the phosphodiester bond at the 3' AG boundary of the intron. This second cut completely releases the intron lariat, which is quickly targeted and degraded, while simultaneously ligating Exon 1 and Exon 2 together into a continuous, unbroken protein-coding sequence.

Visual Connection

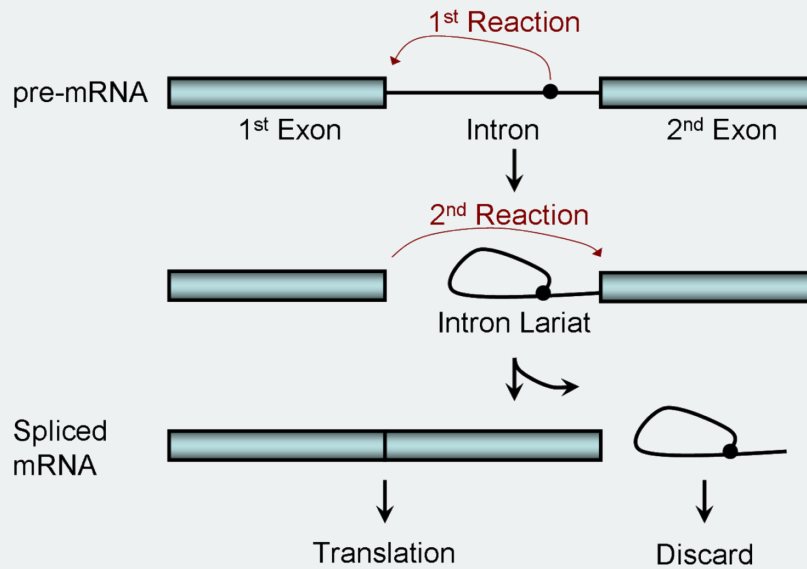


Figure 3.10 Pre-mRNA splicing involves the precise removal of introns from the primary RNA transcript. The splicing process is catalyzed by spliceosomes that are composed of proteins and RNA molecules called small nuclear RNAs (snRNAs). Spliceosomes recognize sequences at the 5' and 3' end of the intron. Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

Note that more than 70 individual introns can be present, and each has to undergo the process of splicing—in addition to 5' capping and the addition of a poly-A tail—just to generate a single, translatable mRNA molecule.

Link to Learning

Link to Learning

See how introns are removed during RNA splicing [at this website](#).

Alternative Splicing

While the precise removal of introns is necessary, the splicing mechanism also provides eukaryotic cells with an advantage: **alternative splicing**. Rather than always joining exons together in a single, fixed linear sequence, the spliceosome can selectively include or skip specific exons under different cellular conditions or in different tissues. By mixing and matching distinct combinations of exons from a single pre-mRNA transcript, a cell can generate multiple unique mature mRNA variants. These variants are subsequently translated into a diverse family of related proteins, known as **protein isoforms**.

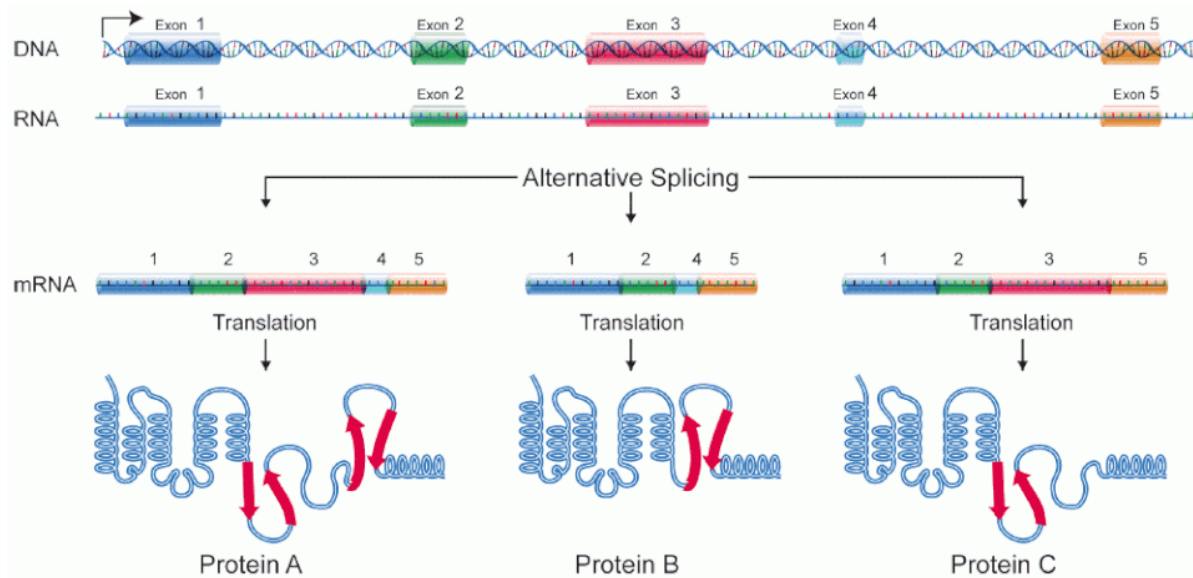


Figure 3.11 Alternative splicing can selectively exclude exons from the mature mRNA, leading to different forms of the protein that may have different functions.

Processing of tRNAs and rRNAs

While messenger RNAs carry the instructions to build proteins, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) are non-coding structural molecules that drive the translation machinery. Unlike mRNAs, these transcripts are never translated into polypeptide chains. They must undergo a series of chemical and structural modifications to achieve their active forms. In eukaryotes, **pre-rRNAs** are transcribed, heavily processed, and assembled alongside ribosomal proteins inside the nucleolus. Concurrently, **pre-tRNAs** are transcribed and trimmed within the nucleus before being exported to the cytoplasm, where they are enzymatically charged with free amino acids to support active protein synthesis.

Ribosomal RNA Processing and Assembly

In both prokaryotes and eukaryotes, multiple individual rRNAs are initially synthesized as a single, massive polycistronic precursor molecule. Specific endonucleases and exonucleases must then precisely cleave and trim this precursor transcript into individual structural subunits. In eukaryotic cells, a single 45S pre-rRNA transcript is synthesized by RNA Polymerase I and sequentially cleaved into the mature **18S** (which forms the core of the small 40S ribosomal subunit), **5.8S**, and **28S** rRNAs (which combine with a separately transcribed 5S rRNA to form the large 60S ribosomal subunit).

Prior to and during these cleavage events, the pre-rRNA molecule undergoes chemical modification, including the addition of protective methyl functional groups ($-\text{CH}_3$) to the 2'-hydroxyl groups of specific ribose sugars, as well as the conversion of standard uridines to **pseudouridines**. These modifications prevent enzymatic degradation and ensure proper structural folding. Rather than finding these modification targets at random, the modifying enzymes are physically guided to the exact target nucleotides by a class of small helper molecules known as **small nucleolar RNAs (snoRNAs)**. Once fully processed, these mature rRNAs account for approximately 50 to 60 percent of the total mass of a functional ribosome, acting as the core catalytic ribozyme that forms peptide bonds during translation.

Transfer RNA Processing and Structural Maturation

Pre-tRNA transcripts undergo a maturation pathway before they can function in the cytoplasm. First, the extended 5' leader sequence of the precursor tRNA is cleaved by a specialized endonuclease called **RNase P**, a historically significant molecule celebrated as one of the first discovered ribozymes because its catalytic activity is driven by its RNA component rather than its protein structure. Next, the 3' trailer sequence is trimmed away by specific exonucleases. Because many raw genomic tRNA segments do not encode the terminal binding sequence for amino acids, an enzyme called **tRNA nucleotidyltransferase** must then template-independently append a conserved trinucleotide **CCA sequence to the extreme 3' hydroxyl end** of every tRNA

molecule.

Similar to rRNAs, pre-tRNAs undergo heavy base modifications, including methylation and chemical rearrangements, which allow the single strand to fold into its precise three-dimensional tertiary conformation. This structure is held rigid by local regions of intramolecular base pairing stabilized by internal hydrogen bonds, giving its characteristic "cloverleaf" secondary shape that folds into an L-shaped tertiary structure. This structural configuration is functional: it positions the newly synthesized **3' CCA amino acid attachment site** at one end of the molecule, and the critical **anticodon loop** at the opposite end. The anticodon consists of a unique three-nucleotide sequence that reads the genetic code by physically interacting with complementary three-nucleotide mRNA codons through antiparallel hydrogen bonding, ensuring that the correct amino acid is faithfully deposited into the growing protein chain.

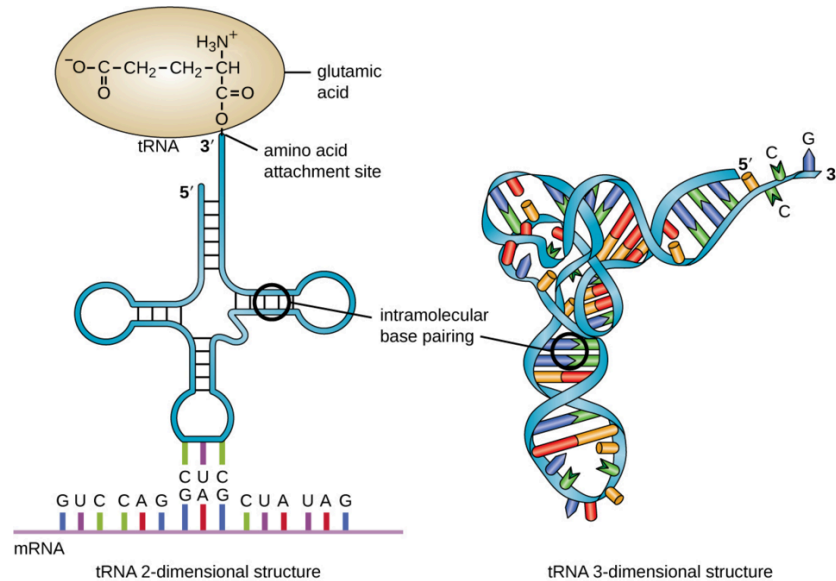


Figure 3.12 Structure of a tRNA showing its interaction with the mRNA and covalently bound (charged) amino acid.

Key Terms

Pre-mRNA (Primary Transcript)

The initial RNA molecule produced by RNA Polymerase II before processing.

RNA Processing

The series of modifications that convert pre-mRNA into mature mRNA.

Mature mRNA

A fully processed messenger RNA that is ready for export from the nucleus and translation.

5' Cap

A modified guanine nucleotide added to the 5' end of eukaryotic mRNA.

Poly-A Tail

A sequence of adenine nucleotides added to the 3' end of eukaryotic mRNA.

RNA Splicing

The removal of introns and joining of exons to produce mature mRNA.

Intron

A non-coding sequence within a gene that is transcribed but removed during processing.

Exon

A coding sequence within a gene that is preserved during processing and expressed in the final protein product.

Spliceosome

The molecular machine that removes introns

Alternative Splicing

Exons are differentially included or excluded in the mature mRNA, producing multiple distinct protein isoforms from a single gene.

Check Your Understanding

1. What is pre-mRNA?
2. Why can't eukaryotic pre-mRNA be translated immediately like prokaryotic mRNA ?
3. What are the three major steps of RNA processing?
4. How does RNA processing improve mRNA stability?
5. What is alternative splicing, and how does it explain the discrepancy between the number of genes in the human genome (20,000) and the total number of unique proteins produced (100,000+)?

2: MODULE 2

2.1: Mitosis, Meiosis, and the Cell Cycle

- 2.1.1: The Cell Cycle
- 2.1.2: Control of the Cell Cycle
- 2.1.3: Cancer and the Cell Cycle
- 2.1.4: Prokaryotic Cell Division
- 2.1.5: Meiosis
- 2.1.6: Errors in Meiosis

2.2: Translation

- 2.2.1: The Genetic Code
- 2.2.2: Ribosomes and Protein Synthesis

2.3: Gene Expression

- 2.3.1: Regulation of Gene Expression
- 2.3.2: Prokaryotic Gene Regulation
- 2.3.3: Eukaryotic Gene Regulation
- 2.3.4: Cancer and Gene Regulation

2.4: Mutations and DNA Repair

- 2.4.1: DNA Damage and Repair
- 2.4.2: Types of Mutations

2.1: MITOSIS, MEIOSIS, AND THE CELL CYCLE

2.1.1: The Cell Cycle

2.1.2: Control of the Cell Cycle

2.1.3: Cancer and the Cell Cycle

2.1.4: Prokaryotic Cell Division

2.1.5: Meiosis

2.1.6: Errors in Meiosis

This page titled [2.1: Mitosis, Meiosis, and the Cell Cycle](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Jennifer Yeack.

2.1.1: The Cell Cycle

The cell cycle is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and nuclear and cytoplasmic division that ultimately produces two identical (clone) cells. The cell cycle has two major phases: interphase and the mitotic phase (Figure 7.1). During interphase, the cell grows and DNA is replicated. During the mitotic phase, the replicated DNA and cytoplasmic contents are separated, and the cell cytoplasm is typically partitioned by a third process of the cell cycle called cytokinesis.

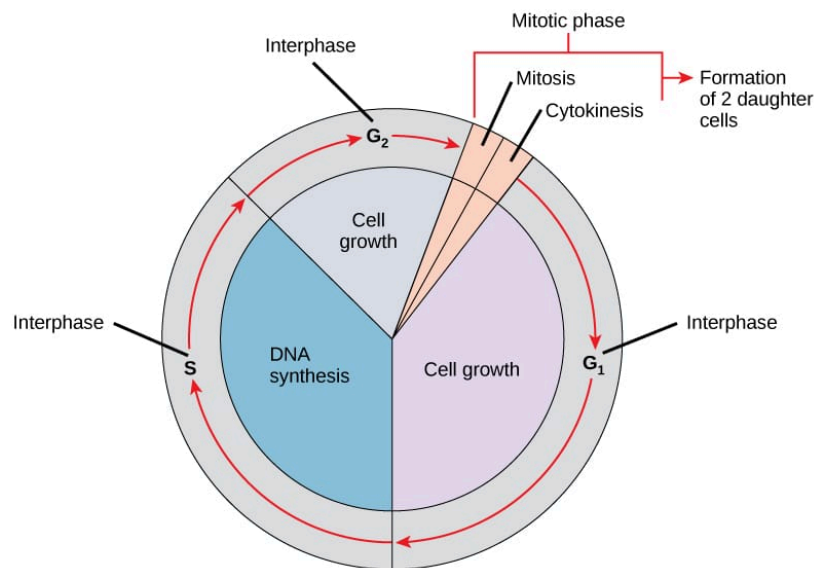


Figure 7.1 The cell cycle in multicellular organisms consists of interphase and the mitotic phase. During interphase, the cell grows and the nuclear DNA is duplicated. Interphase is followed by the mitotic phase. During the mitotic phase, the duplicated chromosomes are segregated and distributed into daughter nuclei. Following mitosis, the cytoplasm is usually divided as well by cytokinesis, resulting in two genetically identical daughter cells.

Interphase

During interphase, the cell undergoes normal growth processes while also preparing for cell division. In order for a cell to move from interphase into the mitotic phase, many internal and external conditions must be met. The three stages of interphase are called *G₁*, *S*, and *G₂*.

G₁ Phase

The first stage of interphase is called the *G₁* phase because, from a microscopic point of view, little change is visible. However, during the *G₁* 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 sufficient energy reserves to complete the task of replicating each chromosome in the nucleus.

S Phase (Synthesis of DNA)

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In the S phase, DNA replication can proceed through the mechanisms that result in the formation of identical pairs of DNA molecules—**sister chromatids**—that are firmly attached at the centromeric region. The centrosome is also 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. For example, roughly at the center of each animal cell, the centrosomes are associated with a pair of rod-like objects, the centrioles, which are positioned

at right angles to each other. Centrioles help organize cell division. We should note, however, that centrioles are not present in the centrosomes of other eukaryotic organisms, such as plants and most fungi.

G₂ Phase

In the G₂ phase, the cell replenishes its energy stores and synthesizes proteins necessary for chromosome manipulation and movement. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic phase. There may be additional cell growth during G₂. 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

The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and move into two new, identical daughter cells. The first portion of the mitotic phase is called karyokinesis, or nuclear division. The second portion of the mitotic phase is called cytokinesis—the physical separation of the cytoplasmic components into the two daughter cells.

Karyokinesis (Mitosis)

Mitosis is divided into a series of phases—prophase, prometaphase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus.

Prophase (the “first phase”): the nuclear envelope starts to dissociate, and the membranous organelles fragment and disperse toward the periphery of the cell. The nucleolus disappears as well. Microtubules that will form the *mitotic spindle* extend between the centrosomes, *pushing them farther apart* as the microtubule fibers lengthen. The sister chromatids begin to coil more tightly with the aid of **condensin proteins** and now become visible under a light microscope.

Prometaphase (the “first change phase”): Many processes that began in prophase continue to advance. Chromosomes become even more condensed and discrete. Each sister chromatid develops a protein structure called a *kinetochore* in its centromeric region. The proteins of the kinetochore attract and bind to the mitotic spindle microtubules. Once a mitotic fiber attaches to a chromosome, the chromosome will be oriented until the kinetochores of sister chromatids face the *opposite poles*. Eventually, all the sister chromatids will be attached via their kinetochores to microtubules from opposing poles.

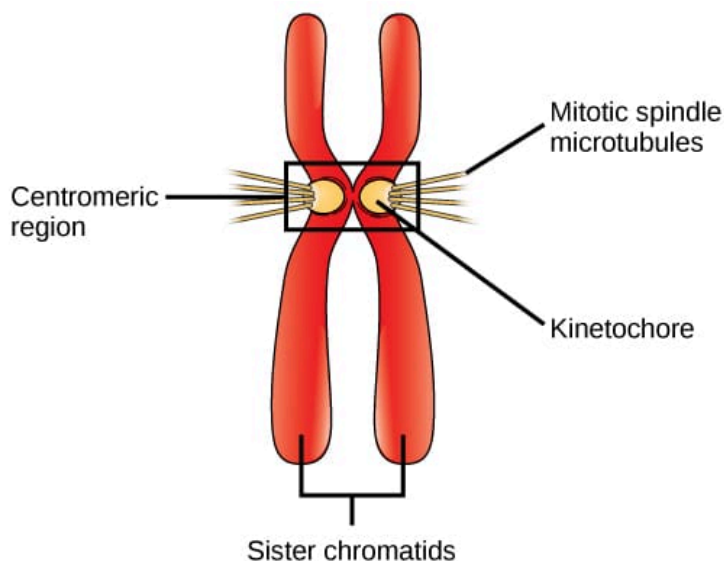


Figure 7.2 During prometaphase, mitotic spindle microtubules from opposite poles attach to each sister chromatid at the kinetochore. In anaphase, the connection between the sister chromatids breaks down, and the microtubules pull the chromosomes toward opposite poles.

Metaphase (the “middle phase”): All the chromosomes are aligned in a plane called the *metaphase plate*, or the equatorial plane, roughly midway between the two poles of the cell. The sister chromatids are still tightly attached to each other by **cohesin proteins**. At this time, the chromosomes are maximally condensed.

Anaphase (“away phase”): The cohesin proteins degrade, and the sister chromatids separate at the centromere. Each chromatid, now a single chromosome, is pulled rapidly toward the opposite poles. The cell becomes visibly elongated (oval shaped) as the polar microtubules slide against each other at the metaphase plate where they overlap.

Telophase (the “two phase”): the chromosomes reach the opposite poles and begin to *decondense* (unravel), relaxing once again into a stretched-out chromatin configuration. The mitotic spindles are depolymerized into tubulin monomers that will be used to assemble cytoskeletal components for each daughter cell. Nuclear envelopes form around the chromosomes, and nucleosomes appear within the nuclear area.

Visual Connection

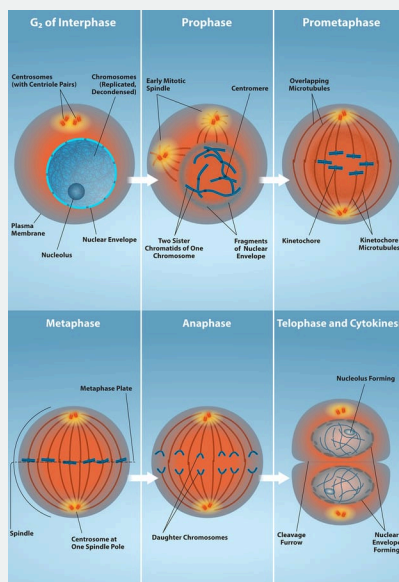


Figure 7.3 G₂ of Interphase – The last stage of interphase is the second gap period, G₂. During this stage, cells grow, replenish energy and synthesize needed macromolecules, such as proteins and lipids.

Mitosis – When G₂ is complete, the cell will enter mitosis. Although there are 5 phases in mitosis, with the exception of the metaphase to anaphase transition, these phases are not discrete and happen as a continuous process. Prophase is the first stage in mitosis. The nuclear envelope begins to break down and chromosomes condense and are now visible. Spindle fibers start to appear and centrosomes begin to move towards opposite poles.

Prometaphase – Chromosomes continue to condense and are more visible. Kinetochore appear at the centromere and kinetochore microtubules attach. Centrosomes continue to move towards opposite poles.

Metaphase – The mitotic spindle is fully developed and centrosomes are at opposite poles. Chromosomes are aligned at the “equatorial plate”, and each sister chromatid rests on one side of the plate, with spindle fibers attached to them.

Anaphase – Sister chromatids are pulled apart by spindle fibers and are separated from each other. Each chromatid is now a chromosome.

Telophase and Cytokinesis – Chromosomes arrive at opposite poles and start to decondense and become less visible. The nuclear envelope reassembles and begins to surround each new set of chromosomes. The mitotic spindle assembly breaks down and the division of the cytoplasm begins via cytokinesis. This physical separation into two cells are remarkably different processes in plant and animal cells.

Credit: Rao, A., Hawkins, A. and Fletcher, S. Department of Biology, Texas A&M University.

Which of the following is the correct order of events in mitosis?

- Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus reforms and the cell divides. Cohesin proteins break down and the sister chromatids separate.
- The kinetochore becomes attached to the mitotic spindle. Cohesin proteins break down and the sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus reforms and the cell divides.

- c. The kinetochore becomes attached to the cohesin proteins. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus reforms and the cell divides.
- d. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. Cohesin proteins break down and the sister chromatids separate. The nucleus reforms and the cell divides.

Cytokinesis

Cytokinesis is sometimes viewed as the second main stage of the mitotic phase, during which cell division is completed via the physical separation of the cytoplasmic components into two daughter cells. However, cytokinesis can also be viewed as a separate phase, which may or may not take place following mitosis. If cytokinesis does take place, cell division is not complete until the cell components have been apportioned and completely separated into the 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 animal cells, cytokinesis typically starts during late 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 is called the cleavage furrow. The furrow deepens as the actin ring contracts, and eventually the membrane is cleaved in two (Figure 7.4).

In plant cells, 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 into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles are transported on microtubules to form a *phragmoplast* (a vesicular structure) at the metaphase plate. There, the vesicles fuse and coalesce 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 walls at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall. The Golgi membranes become parts of the plasma membrane on either side of the new cell wall (Figure 7.4).

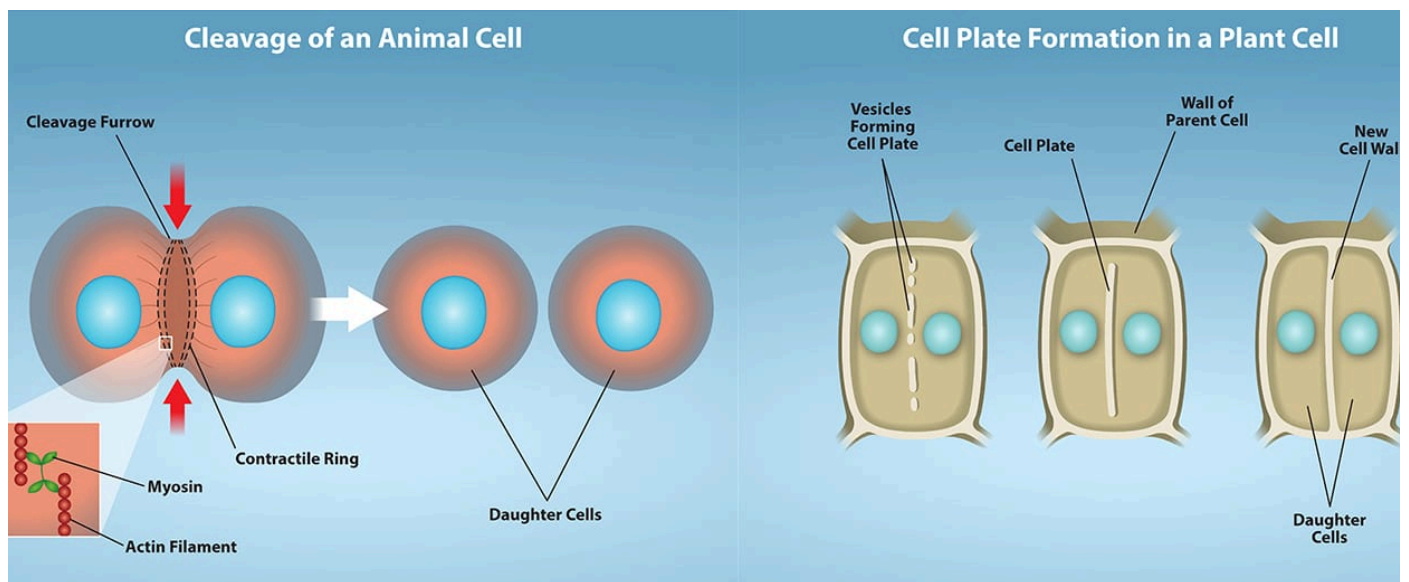


Figure 7.4 Animal cell cleavage - Contraction of actin filaments by myosin pull the “equator” of the cell causing an invagination of the cell (a fissure called a cleavage furrow). The cleavage furrow gets continuously deeper until it eventually divides the cell into two new, independent daughter cells.

Plant cell wall formation - A mixture of enzymes, proteins, and glucose molecules are transported via vesicles to the center of the cell. These vesicles continuously build upon each other until a completely new cell wall has emerged and two new cells are formed, independent of one another.

Credit: Rao, A., Ryan, K., Hawkins, A. and Fletcher, S. Department of Biology, Texas A&M University.

G₀ Phase

Not all cells adhere to the classic cell-cycle pattern in which a newly formed daughter cell immediately enters the preparatory phases of interphase, closely followed by the mitotic phase, and cytokinesis. Cells in **G₀** phase are not actively preparing to

divide. The cell is in a quiescent (inactive) stage that occurs when cells exit the cell cycle. Some cells enter G_0 temporarily due to environmental conditions such as availability of nutrients, or stimulation by growth factors. The cell will remain in this phase until conditions improve or 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.

Key Terms

Cell Cycle

The ordered sequence of events through which a cell grows, replicates its DNA, and divides to produce two daughter cells.

Interphase

The longest phase of the cell cycle during which the cell grows, carries out normal cellular functions, replicates its DNA, and prepares for division.

G_1 Phase

The first stage of interphase in which the cell grows, synthesizes proteins and organelles, and prepares for DNA replication.

S Phase

The stage of interphase during which DNA is replicated, producing identical sister chromatids.

G_2 Phase

The final stage of interphase during which the cell grows, synthesizes proteins needed for mitosis, and prepares for cell division.

Mitotic Phase

The phase of the cell cycle consisting of mitosis (nuclear division) and cytokinesis (cytoplasmic division).

Mitosis

The process of nuclear division that separates duplicated chromosomes into two genetically identical nuclei.

Cytokinesis

The division of the cytoplasm that produces two separate daughter cells following mitosis.

Sister Chromatids

Two identical copies of a chromosome produced during DNA replication and joined together at the centromere.

Centromere

The constricted region of a chromosome where sister chromatids are attached and where the kinetochore forms.

Kinetochore

A protein complex assembled at the centromere that serves as the attachment site for spindle microtubules during mitosis.

Mitotic Spindle

A network of microtubules that separates chromosomes during mitosis.

Condensin

A protein complex that helps condense chromosomes during mitosis.

Cohesin

A protein complex that holds sister chromatids together until anaphase.

Prophase

The first stage of mitosis in which chromosomes condense, the nucleolus disappears, and the mitotic spindle begins to form.

Prometaphase

The stage of mitosis in which the nuclear envelope breaks down completely and spindle fibers attach to kinetochores.

Metaphase

The stage of mitosis in which chromosomes align along the metaphase plate.

Metaphase Plate

An imaginary plane located midway between the spindle poles where chromosomes align during metaphase.

Anaphase

The stage of mitosis in which sister chromatids separate and move toward opposite poles of the cell.

Telophase

The final stage of mitosis in which chromosomes arrive at opposite poles, decondense, and new nuclear envelopes form.

G₀ Phase

A resting or quiescent state in which cells have exited the cell cycle and are not actively preparing to divide.

Check Your Understanding

1. What is the purpose of the cell cycle?
2. Name the two major phases of the cell cycle.
3. Why must the events of the cell cycle occur in a specific order?
4. What is the primary purpose of interphase?
5. Why is DNA replicated during the S phase?
6. What cellular changes occur during the G₂ phase?
7. At what point are sister chromatids first formed?
8. Place the stages of mitosis in the correct order.
9. What major events occur during prophase?
10. Why is chromosome alignment at the metaphase plate important?
11. What event marks the beginning of anaphase?
12. What happens to chromosomes during telophase?
13. What roles do cohesin and condensin proteins play during mitosis?
14. Compare cytokinesis in animal cells and plant cells.
15. What is the G₀ phase?
16. Why do some cells enter G₀ only temporarily?

2.1.2: Control of the Cell Cycle

The length of the cell cycle is highly variable, even within the cells of a single 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, and 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 rapidly dividing mammalian cells are grown in a culture (outside the body under optimal growing conditions), the length of the cell cycle is about 24 hours. In rapidly dividing human cells with a 24-hour cell cycle, the G_1 phase lasts approximately nine hours, the S phase lasts 10 hours, the G_2 phase lasts about four and one-half hours, and the M phase lasts approximately one-half hour. By comparison, in fertilized eggs (and early embryos) of fruit flies, the cell cycle is completed in about eight minutes. This is because the nucleus of the fertilized egg divides many times by mitosis but does not go through cytokinesis until a multinucleate “zygote” has been produced, with many nuclei located along the periphery of the cell membrane, thereby shortening the time of the cell division cycle. The timing of events in the cell cycle of both “invertebrates” and “vertebrates” is controlled by mechanisms that are both internal and external to the cell.

Regulation of the Cell Cycle by External Events

Both the initiation and inhibition of cell division are triggered by events external to the cell when it is about to begin the replication process. An event may be as simple as the death of nearby cells or as sweeping as the release of growth-promoting hormones, such as human growth hormone (HGH or hGH). A lack of HGH can *inhibit* cell division, resulting in dwarfism, whereas too much HGH can result in gigantism. Crowding of cells can also inhibit cell division. In contrast, a factor that can initiate cell division is the size of the cell: As a cell grows, it becomes physiologically inefficient due to its decreasing surface-to-volume ratio. The solution to this problem is to divide. Whatever the source of the message, the cell receives the signal, and a series of events within the cell allows it to proceed into interphase. Moving forward from this initiation point, every parameter required during each cell cycle phase must be met or the cycle cannot progress.

Regulation at Internal Checkpoints

It is essential that the daughter cells produced 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 an abnormal cell. To prevent a compromised cell from continuing to divide, there are internal control mechanisms that operate at three main **cell-cycle checkpoints**: A checkpoint is one of several points in the eukaryotic cell cycle at which the progression of a cell to the next stage in the cycle can be halted until conditions are favorable. These checkpoints occur near the end of G_1 , at the G_2/M transition, and during metaphase (Figure 7.5).

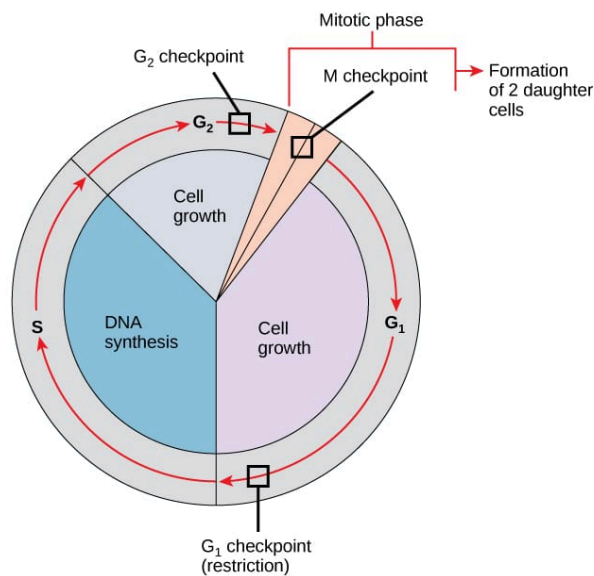


Figure 7.5 The cell cycle is controlled at three checkpoints. The integrity of the DNA is assessed at the G₁ checkpoint. Proper chromosome duplication is assessed at the G₂ checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

The G₁ Checkpoint

The **G₁ checkpoint** determines whether all conditions are favorable for cell division to proceed. The G₁ checkpoint, also called the restriction point (in yeast), is a point at which the cell irreversibly commits to the cell division process. External influences, such as growth factors, play a large role in carrying the cell past the G₁ checkpoint. In addition to adequate reserves and cell size, there is a check for genomic DNA damage at the G₁ checkpoint. A cell that does not meet all the requirements will not be allowed to progress into the S phase. The cell can halt the cycle and attempt to remedy the problematic condition, or the cell can advance into G₀ and await further signals when conditions improve.

The G₂ Checkpoint

The **G₂ checkpoint** bars entry into the mitotic phase if certain conditions are not met. As at the G₁ checkpoint, cell size and protein reserves are assessed. However, the most important role of the G₂ checkpoint is to ensure that *all of the chromosomes have been replicated and that the replicated DNA is not damaged*. If the checkpoint mechanisms detect problems with the DNA, the cell cycle is halted, and the cell attempts to either complete DNA replication or repair the damaged DNA.

The M Checkpoint

The **M checkpoint** occurs near the end of the metaphase stage of karyokinesis. The M checkpoint is also known as the spindle checkpoint, because it determines whether 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 at least two spindle fibers arising from opposite poles of the cell.

Regulator Molecules of the Cell Cycle

In addition to the internally controlled checkpoints, there are two groups of intracellular molecules that regulate the cell cycle. These regulatory molecules either promote progress of the cell to the next phase (positive regulation) or halt the cycle (negative regulation). Regulator molecules may act individually, or they can influence the activity or production of other regulatory proteins. Therefore, the failure of a single regulator may have almost no effect on the cell cycle, especially if more than one mechanism controls the same event. However, the effect of a deficient or non-functioning regulator can be wide-ranging and possibly fatal to the cell if multiple processes are affected.

Positive Regulation of the Cell Cycle

Two groups of proteins, called cyclins and cyclin-dependent kinases (Cdks), are termed **positive regulators**. They are responsible for the progress of the cell through the various checkpoints. The levels of the four cyclin proteins fluctuate throughout the cell cycle in a predictable pattern (Figure 7.6). Increases in the concentration of cyclin proteins are triggered by both external and internal signals. After the cell moves to the next stage of the cell cycle, the cyclins that were active in the previous stage are degraded by cytoplasmic enzymes.

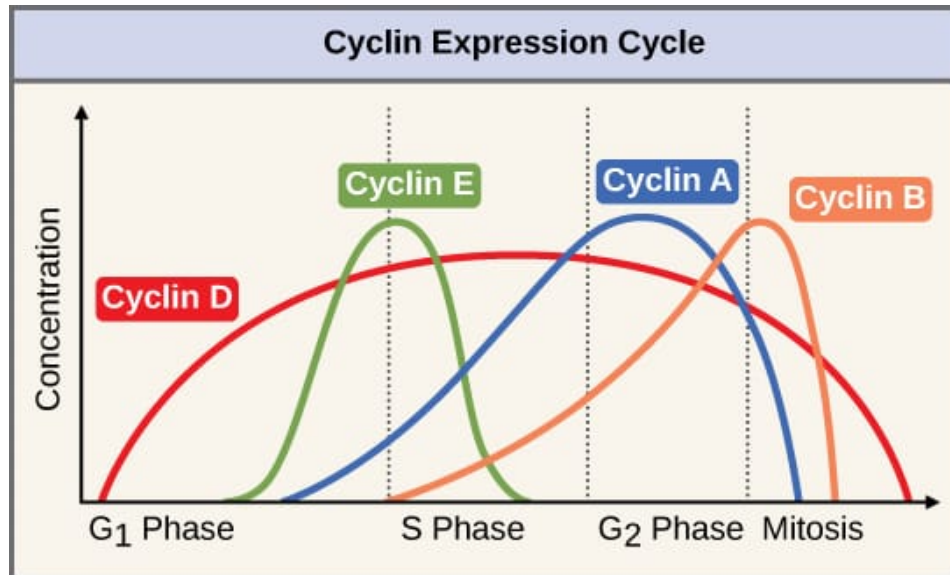


Figure 7.6 The concentrations of cyclin proteins change throughout the cell cycle. There is a direct correlation between cyclin accumulation and the three major cell-cycle checkpoints. Also note the sharp decline of cyclin levels following each checkpoint (the transition between phases of the cell cycle), as cyclin is degraded by cytoplasmic enzymes. (credit: modification of work by "WikiMiMa"/Wikimedia Commons)

Cyclins regulate the cell cycle only when they are tightly bound to Cdks. To be fully active, the Cdk/cyclin complex must also be phosphorylated in specific locations to activate the complex. Like all kinases, Cdks are enzymes (*kinases*) that in turn phosphorylate other proteins. Phosphorylation activates the protein by changing its shape. The proteins phosphorylated by Cdks are involved in advancing the cell to the next phase. (Figure 7.7). The levels of Cdk proteins are relatively stable throughout the cell cycle; however, the concentrations of cyclin fluctuate and determine when Cdk/cyclin complexes form. The different cyclins and Cdks bind at specific points in the cell cycle and thus regulate different checkpoints.

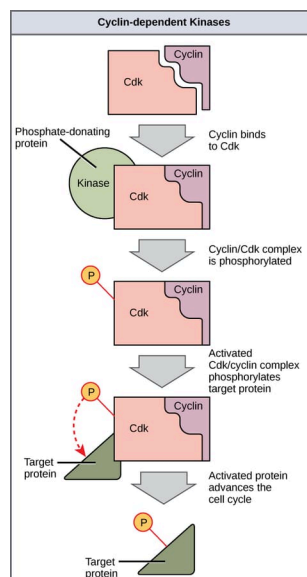


Figure 7.7 *Cyclin-dependent kinases (Cdks)* are protein kinases that, when fully activated, can phosphorylate and thus activate other proteins that advance the cell cycle past a checkpoint. To become fully activated, a Cdk must bind to a cyclin protein and then be phosphorylated by another kinase.

Because the cyclic fluctuations of cyclin levels are largely based on the *timing of the cell cycle* and not on specific events, regulation of the cell cycle usually occurs by either the Cdk molecules alone or the Cdk/cyclin complexes. Without a specific concentration of fully activated cyclin/Cdk complexes, the cell cycle cannot proceed through the checkpoints.

Although the cyclins are the main regulatory molecules that determine the forward momentum of the cell cycle, there are several other mechanisms that fine-tune the progress of the cycle with negative, rather than positive, effects. These mechanisms essentially block the progression of the cell cycle until problematic conditions are resolved. Molecules that prevent the full activation of Cdks are called Cdk inhibitors. Many of these inhibitor molecules directly or indirectly monitor a particular cell-cycle event. The block placed on Cdks by inhibitor molecules will not be removed until the specific event that the inhibitor monitors is completed.

Negative Regulation of the Cell Cycle

The second group of cell-cycle regulatory molecules are **negative regulators**, which stop the cell cycle. Remember that in positive regulation, active molecules cause the cycle to progress. The best understood negative regulatory molecules are retinoblastoma protein (Rb), p53, and p21. Retinoblastoma proteins are a group of *tumor-suppressor proteins* common in many cells. We should note here that the 53 and 21 designations refer to the functional molecular masses of the proteins (p) in kilodaltons (a dalton is equal to an *atomic mass unit*, which is equal to one proton or one neutron or 1 g/mol). Much of what is known about cell-cycle regulation comes from research conducted with cells that have *lost regulatory control*. All three of these regulatory proteins were discovered to be damaged or non-functional in cells that had begun to replicate uncontrollably (i.e., became cancerous). In each case, the main cause of the unchecked progress through the cell cycle was a faulty copy of the regulatory protein.

Rb, p53, and p21 act primarily at the G₁ checkpoint. p53 is a multi-functional protein that has a major impact on the commitment of a cell to division because it acts when there is damaged DNA in cells that are undergoing the preparatory processes during G₁. If damaged DNA is detected, p53 halts the cell cycle and then recruits specific enzymes to repair the DNA. If the DNA cannot be repaired, p53 can trigger apoptosis, or cell suicide, to prevent the duplication of damaged chromosomes. As p53 levels rise, the production of p21 is triggered. p21 enforces the halt in the cycle dictated by p53 by binding to and inhibiting the activity of the Cdk/cyclin complexes. As a cell is exposed to more stress, higher levels of p53 and p21 accumulate, making it less likely that the cell will move into the S phase.

Rb, which largely monitors cell size, exerts its regulatory influence on other positive regulator proteins. In the *active*, dephosphorylated state, Rb binds to proteins called *transcription factors*, most commonly, E2F (Figure 7.8). Transcription factors “turn on” specific genes, allowing the production of proteins encoded by that gene. When Rb is bound to E2F, production of proteins necessary for the G₁/S transition is blocked. As the cell increases in size, Rb is slowly phosphorylated until it becomes *inactivated*. Rb releases E2F, which can now turn on the gene that produces the transition protein, and this particular block is removed. For the cell to move past each of the checkpoints, all positive regulators must be “turned on,” and all negative regulators must be “turned off.”

Visual Connection

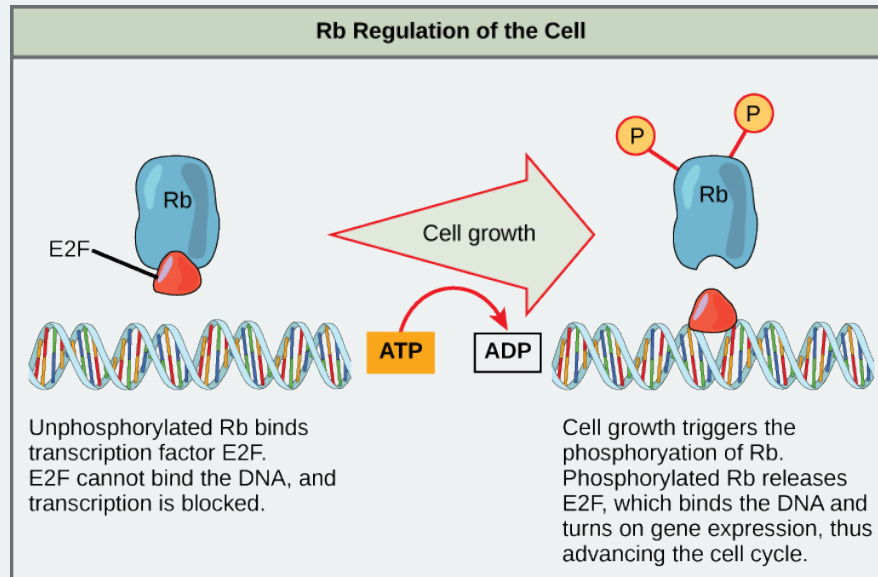


Figure 7.8 Rb halts the cell cycle and releases its hold in response to cell growth.

Rb and other proteins that negatively regulate the cell cycle are sometimes called tumor suppressors. Why do you think the name tumor suppressor might be appropriate for these proteins?

Key Terms

Cell Cycle Regulation

The mechanisms that control the timing and progression of the cell cycle to ensure accurate cell growth, DNA replication, and cell division.

Cell Cycle Checkpoint

A control point in the cell cycle where the cell assesses whether conditions are favorable before progressing to the next stage.

G₁ Checkpoint (Restriction Point)

The checkpoint near the end of G₁ phase where the cell determines whether conditions are favorable for DNA replication. Once a cell passes this checkpoint, it is generally committed to completing the cell cycle.

G₂ Checkpoint

The checkpoint that verifies DNA replication has been completed accurately and that any DNA damage has been repaired before the cell enters mitosis.

M Checkpoint (Spindle Checkpoint)

The checkpoint during metaphase that ensures all chromosomes are properly attached to spindle microtubules before sister chromatids separate.

Positive Regulator

A molecule that promotes progression through the cell cycle.

Negative Regulator

A molecule that slows or halts the cell cycle when errors or unfavorable conditions are detected.

Retinoblastoma Protein (Rb)

A tumor suppressor protein that prevents progression through the G₁ checkpoint by inhibiting transcription factors required for DNA replication.

E2F

A transcription factor that activates genes necessary for entry into the S phase. It is inhibited by active Rb protein.

p53

A tumor suppressor protein that monitors DNA damage, halts the cell cycle, activates DNA repair, or triggers apoptosis if the damage cannot be repaired.

p21

A protein produced in response to p53 activation that inhibits cyclin-Cdk complexes and enforces cell cycle arrest.

Apoptosis

A controlled process of programmed cell death that eliminates damaged or unnecessary cells.

Check Your Understanding

1. Why is regulation of the cell cycle essential for multicellular organisms?
2. How can external signals influence whether a cell enters the cell cycle?
3. What is the purpose of a cell cycle checkpoint?
4. Compare the functions of the G₁, G₂, and M checkpoints.
5. What conditions must be satisfied before a cell can pass the G₂ checkpoint?
6. Why is the M checkpoint particularly important for maintaining chromosome number?
7. Compare positive regulators and negative regulators of the cell cycle.
8. What role does p53 play when DNA damage is detected?
9. How does p21 help enforce cell cycle arrest?
10. Why are Rb and p53 classified as tumor suppressor proteins?
11. What happens if DNA damage cannot be repaired?

2.1.3: Cancer and the Cell Cycle

Cancer comprises many different diseases caused by a common mechanism: uncontrolled cell growth. Despite the redundancy and overlapping levels of cell-cycle control, errors do 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 changes to the DNA nucleotide sequence occur within a coding portion of a gene and are not corrected, a gene mutation results. All cancers start when a gene mutation gives rise to a faulty protein that plays a key role in cell reproduction.

The change in the cell that results from the malformed protein may be minor: perhaps a slight delay in the binding of Cdk to cyclin or an Rb protein that detaches from its target DNA while still phosphorylated. Even minor mistakes, however, may allow subsequent mistakes to occur more readily. Over and over, small uncorrected errors are passed from the parent cell to the daughter cells and amplified as each generation 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 (“-oma”) can result.

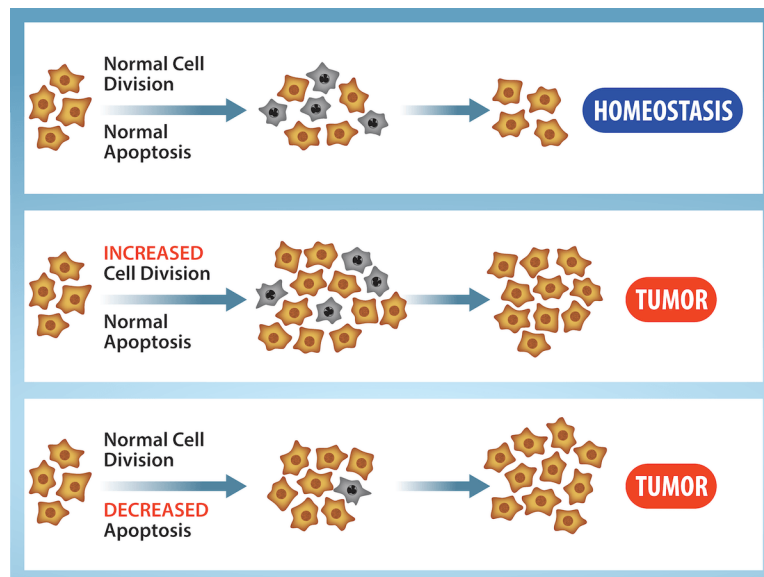


Figure 7.9 Cell Division and Apoptosis. In an adult organism, normal cell division is balanced by apoptosis (programmed cell death) to maintain a constant cell number in homeostasis. Either an increase in cell division or a decrease in apoptosis leads an increase in the number of cells and tumor formation. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

Proto-oncogenes

The genes that code for the positive cell-cycle regulators are called proto-oncogenes. Proto-oncogenes are normal genes that, when mutated in certain ways, 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 to be activated without being partnered with cyclin 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 undergo further cell divisions, the mutation would not be propagated and no harm would come to the organism. However, if the atypical daughter cells are able to undergo further cell divisions, subsequent generations of cells may accumulate even more mutations, some possibly in additional genes that regulate the cell cycle.

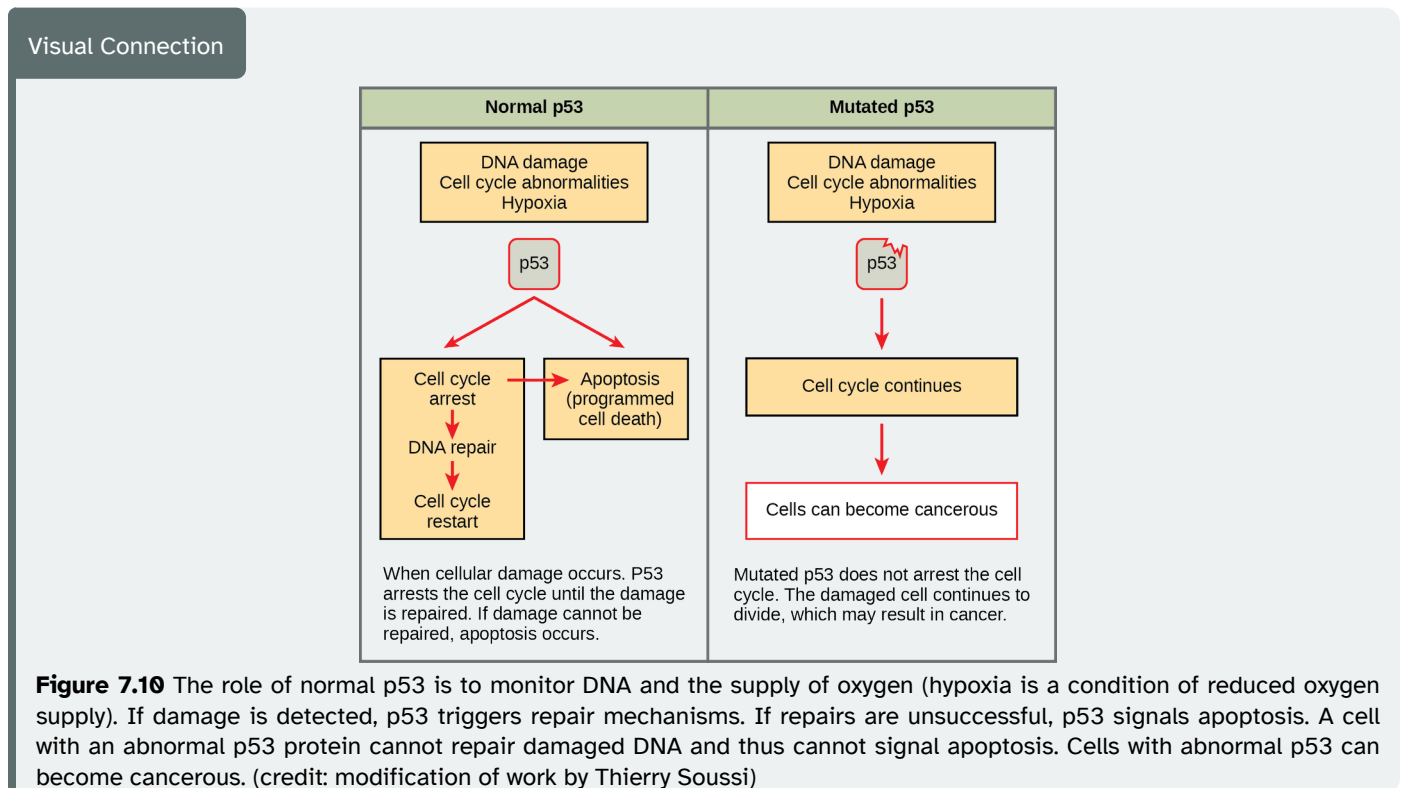
The Cdk gene in the above example is only one of many genes that are considered proto-oncogenes. In addition to the cell-cycle

regulatory proteins, any protein that influences the cycle can be altered in such a way as to override cell-cycle checkpoints. An oncogene is any gene that, when altered, leads to an increase in the rate of cell-cycle progression.

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 segments of DNA that code for negative regulator proteins, the type of regulators that, when activated, can prevent the cell from undergoing uncontrolled division. The collective function of the best-understood tumor suppressor gene proteins, Rb, p53, and p21, is to put up a roadblock to cell-cycle progression 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. Tumor suppressors are similar to brakes in a vehicle: Malfunctioning brakes can contribute to a car crash!

Mutated p53 genes have been identified in more than 50 percent of all human tumor cells. This discovery is not surprising in light of the multiple roles that the p53 protein plays at the G₁ checkpoint. A cell with a faulty p53 may fail to detect errors present in the genomic DNA (Figure 10.16). Even if a partially functional p53 does identify the mutations, it may no longer be able to signal the necessary DNA repair enzymes. Either way, damaged DNA will remain uncorrected. At this point, a functional p53 will deem the cell unsalvageable and trigger programmed cell death (apoptosis). The damaged version of p53 found in cancer cells, however, cannot trigger apoptosis.



Human papillomavirus can cause cervical cancer. The virus encodes E6, a protein that binds p53. Based on this fact and what you know about p53, what effect do you think E6 binding has on p53 activity?

- E6 activates p53
- E6 inactivates p53
- E6 mutates p53
- E6 binding marks p53 for degradation

The loss of p53 function has other repercussions for the cell cycle. Mutated p53 might lose its ability to trigger p21 production. Without adequate levels of p21, there is no effective block on Cdk activation. Essentially, without a fully functional p53, the G₁ checkpoint is severely compromised and the cell proceeds directly from G₁ to S regardless of internal and external conditions. At the completion of this shortened cell cycle, two daughter cells are produced that have inherited the mutated p53 gene. Given

the non-optimal conditions under which the parent cell reproduced, it is likely that the daughter cells will have acquired other mutations in addition to the faulty tumor-suppressor gene. Cells such as these daughter cells quickly accumulate both oncogenes and non-functional tumor-suppressor genes. Again, the result is tumor growth.

2.1.4: Prokaryotic Cell Division

Prokaryotes, such as bacteria, produce daughter cells 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 cloned offspring, certain 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 cellular machinery to sustain life. As we've seen with bacterial cells, the genome consists of a single, circular DNA chromosome; therefore, the process of cell division is simplified. Karyokinesis is unnecessary because there is no true nucleus and thus no need to direct one copy of the multiple chromosomes into each daughter cell. This type of cell division is called binary fission.

Binary Fission

The cell division process in prokaryotes is a less complicated and much more rapid process than cell division in eukaryotes. As a review of the general information on cell division we discussed at the beginning of this chapter, recall that the single, circular DNA chromosome of bacteria occupies a specific location, the nucleoid region, within the cell. The packing proteins of bacteria are, however, related to the cohesin and condensin proteins involved in the chromosome compaction of eukaryotes.

The bacterial chromosome is attached to the plasma membrane at about the midpoint of the cell. The starting point of replication, the origin, is close to the binding site of the chromosome to the plasma membrane (Figure 7.11). Replication of the DNA is bidirectional, moving away from the origin on both strands of the loop simultaneously. As the new double strands are formed, each origin point moves away from the cell wall attachment toward the 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. The formation of a ring composed of repeating units of a protein called FtsZ (short for “filamenting temperature-sensitive mutant Z”) directs the partition between the nucleoids. 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. A septum is formed between the daughter nucleoids, extending gradually from the periphery toward the center of the cell. When the new cell walls are in place, the daughter cells separate.

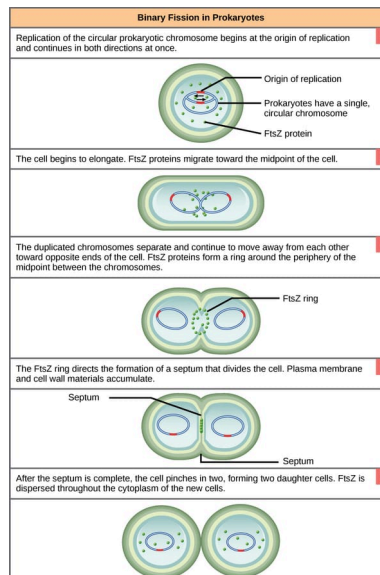


Figure 7.11 These images show the steps of binary fission in prokaryotes. (credit: modification of work by “Mcstrother”/Wikimedia Commons)

Evolution Connection

Evolution Connection

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 karyokinesis 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 which make up the mitotic spindle fibers that are necessary for eukaryotic nuclear division. FtsZ proteins can form filaments, rings, and other three-dimensional structures that resemble the way tubulin forms microtubules, centrioles, and various cytoskeletal 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 considered to be homologous structures derived from common evolutionary origins. In this example, FtsZ is the ancestor protein to tubulin (an evolutionarily derived protein). While both proteins are found in extant organisms, tubulin function has evolved and diversified tremendously since evolving from its FtsZ prokaryotic origin. A survey of mitotic assembly components found in present-day unicellular eukaryotes reveals crucial intermediary steps to the complex membrane-enclosed genomes of multicellular eukaryotes (Table 7.1).

Cell Division Apparatus among Various Organisms

	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 wrapped around histones 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, which 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 centrosomes. The nuclear envelope dissolves. Chromosomes attach to the mitotic spindle, which separates the chromosomes and elongates the cell.	Microfilaments form a cleavage furrow that pinches the cell in two.

Table 7.1

2.1.5: 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 G₁, S, and G₂ phases, which are nearly identical to the phases preceding mitosis. The G₁ 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 G₂ 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 2.1.5.1).

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 7.12) 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.

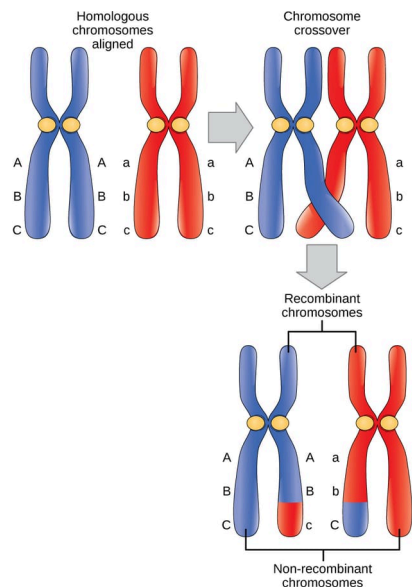


Figure 7.12 In this illustration of the effects of crossing over, the blue chromosome came from the individual's father and the red chromosome came from the individual's mother. Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between 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 7.13).

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.

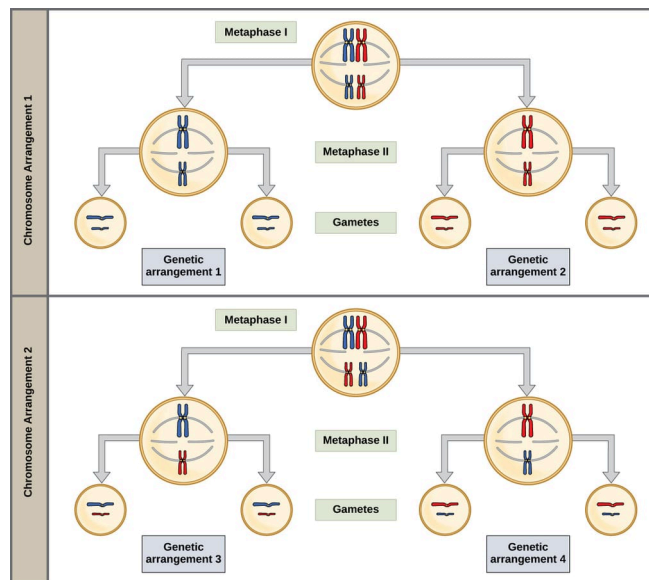


Figure 7.13: To demonstrate random, independent assortment at metaphase I, consider a cell with $n = 2$. In this case, there are two possible arrangements at the equatorial plane in metaphase I, as shown in the upper cell of each panel. These two possible orientations lead to the production of genetically 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 (Figure 7.14).

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 that lacks an S phase, before entering meiosis II. Chromosomes are not duplicated during this time. 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. 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.

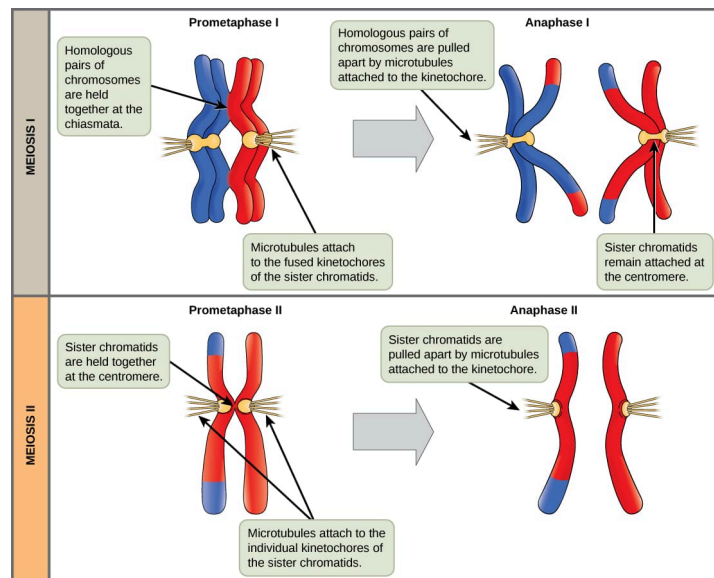


Figure 7.14: In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to individual kinetochores of sister chromatids. In anaphase II, the sister chromatids are separated.

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 (Figure 7.15).

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.

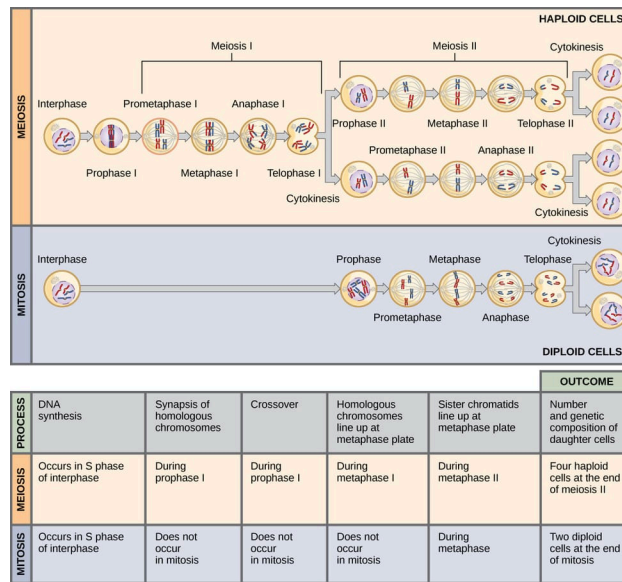


Figure 7.15: Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

CONCEPT IN ACTION



For an animation comparing mitosis and meiosis, go to [this website](#).

Key Terms

Sexual Reproduction

A mode of reproduction in which two haploid gametes fuse during fertilization to produce a diploid offspring.

Fertilization

The fusion of two haploid gametes to form a diploid zygote.

Ploidy

The number of complete sets of chromosomes present in a cell.

Haploid (n)

A cell containing one complete set of chromosomes.

Diploid (2n)

A cell containing two complete sets of chromosomes, one inherited from each parent.

Somatic Cell

A nonreproductive body cell that is typically diploid.

Gamete

A haploid reproductive cell, such as a sperm or egg, that participates in fertilization.

Homologous Chromosomes

A pair of chromosomes that carry the same genes in the same locations but may contain different alleles.

Meiosis

A specialized type of cell division that reduces the chromosome number by half and produces four genetically unique haploid cells.

Reduction Division

The first meiotic division (Meiosis I), during which homologous chromosomes separate and the chromosome number is reduced from diploid to haploid.

Synapsis

The pairing of homologous chromosomes during prophase I of meiosis.

Tetrad

A structure formed during prophase I consisting of two homologous chromosomes and their four sister chromatids.

Crossing Over

The exchange of DNA segments between nonsister chromatids of homologous chromosomes during prophase I.

Chiasma (plural: Chiasmata)

The visible site where crossing over occurs between homologous chromosomes.

Recombinant Chromosome

A chromosome produced by crossing over that contains DNA from both the maternal and paternal homologs.

Independent Assortment

The random orientation of homologous chromosome pairs during metaphase I, resulting in different combinations of maternal and paternal chromosomes in gametes.

Meiosis I

The first meiotic division during which homologous chromosomes separate, reducing the chromosome number by half.

Meiosis II

The second meiotic division during which sister chromatids separate, producing four haploid daughter cells.

Genetic Variation

Differences in DNA sequences among individuals that contribute to biological diversity.

Check Your Understanding

1. Why is meiosis necessary for sexually reproducing organisms?
2. What is the difference between haploid and diploid cells?
3. Compare somatic cells and gametes.

4. What is synapsis, and when does it occur?
5. What is a tetrad?
6. Explain how crossing over produces recombinant chromosomes.
7. During which stage of meiosis does crossing over occur?
8. Why is crossing over considered an important source of genetic variation?
9. What is independent assortment?
10. During which stage of meiosis does independent assortment occur?
11. How does independent assortment increase genetic diversity?
12. Why is Meiosis I called the reduction division?
13. How does Meiosis II resemble mitosis?
14. What happens during anaphase II?
15. Why is DNA not replicated between Meiosis I and Meiosis II?
16. How many daughter cells are produced at the end of meiosis?
17. Compare the chromosome number of daughter cells produced by mitosis and meiosis.
18. Which process produces genetically identical daughter cells?
19. Which process produces genetically unique daughter cells?
20. Name two events that occur during meiosis but not during mitosis.
21. Why is meiosis essential for maintaining chromosome number across generations?

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [2.1.5: Meiosis](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [7.2: Meiosis](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

2.1.6: Errors in Meiosis

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 7.16).

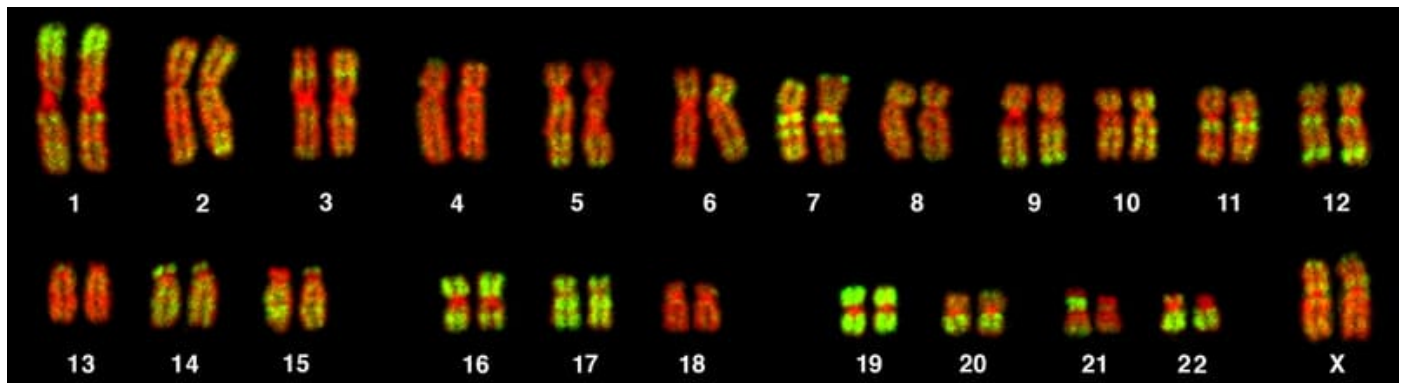


Figure 7.16 This karyogram shows the chromosomes of a female human immune cell during mitosis. (credit: Andreas Bolzer, et al)

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 7.17). 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.

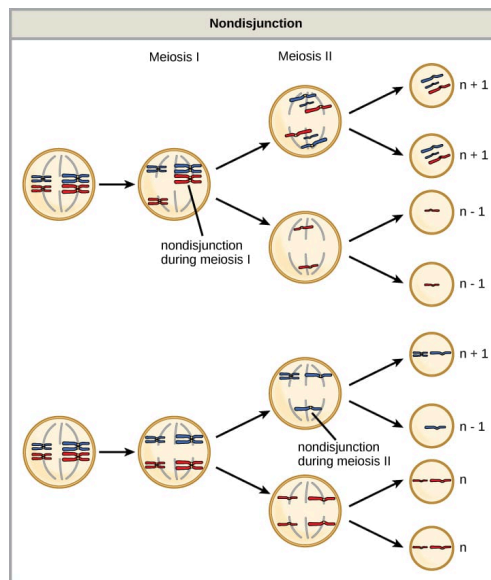


Figure 7.17 Following meiosis, each gamete has one copy of each chromosome. Nondisjunction occurs when homologous chromosomes (meiosis I) or sister chromatids (meiosis II) fail to separate during meiosis.

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** (XX is female and XY is male in humans). 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.

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 7.18). 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 7.18 Embryonic inactivation of one of two different X chromosomes encoding different coat colors gives rise to the tortoiseshell phenotype in cats. (credit: Michael Bodega)

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 triple-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 XO 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. Because these alterations change the physical architecture of the chromosome rather than changing the total number of individual chromosomes, they can have highly varied effects on an organism's development, phenotype, and fertility.

Duplications and Deletions

Duplications and deletions fundamentally alter **gene dosage**, meaning they disrupt the normal balance of cellular proteins by changing the number of copies of specific genes available in the genome. These dosage imbalances often produce offspring that survive but exhibit distinct physical and mental abnormalities. **Deletions** occur when a part of the chromosome is completely lost. The loss of genetic material can eliminate essential genes entirely. For example, **Cri-du-chat** (from the French for “cry of the cat”) is a syndrome associated with severe nervous system abnormalities and identifiable physical features that results from a deletion of most of the small arm (the p arm) of chromosome 5.

Duplications happen when a segment of the chromosome is duplicated. The addition of genetic material introduces extra copies of genes. While this can disrupt metabolic and developmental pathways, gene duplications are also crucially important in

evolutionary biology, as the "extra" copy can mutate and acquire new functions over time without depriving the organism of the original gene's function.

Inversions and Translocations

A chromosome **inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome. Inversions can be classified as either *pericentric* (including the centromere) or *paracentric* (occurring outside the centromere). Unless the break-and-rejoin points directly disrupt a vital structural gene sequence or regulatory element, inversions simply alter the linear orientation of genes. Consequently, they are likely to have much milder phenotypic effects on the individual than aneuploid errors or large deletions.

A **translocation** occurs when a segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome. Translocations can be **reciprocal**, in which segments are cleanly exchanged between two nonhomologous chromosomes without any net loss or gain of genetic material, or **non-reciprocal**, where a segment moves from one chromosome to another without an exchange. Even a balanced reciprocal translocation can have severe clinical effects if it fuses two distinct genes together. A notable example is the **Philadelphia chromosome**, a reciprocal translocation between chromosomes 9 and 22 that creates an abnormal fusion gene responsible for chronic myelogenous leukemia (CML).

EVOLUTION IN ACTION: The Chromosome 18 Inversion

Not 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

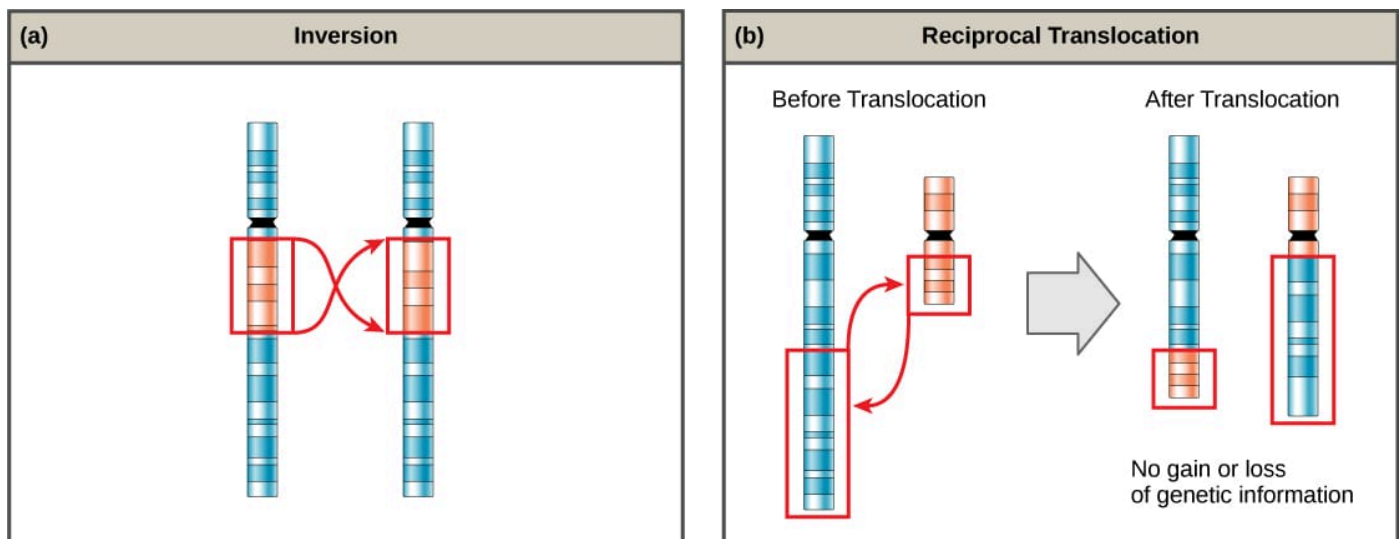


Figure 7.19: An (a) inversion occurs when a chromosome segment breaks from the chromosome, reverses its orientation, and then reattaches in the original position. A (b) reciprocal translocation occurs between two nonhomologous chromosomes and does not cause any genetic information to be lost or duplicated. (credit: modification of work by National Human Genome Research Institute (USA))

Key Terms

Karyotype

The complete set of chromosomes of an individual, including their number, size, shape, and banding pattern.

Nondisjunction

The failure of homologous chromosomes or sister chromatids to separate properly during meiosis or mitosis, resulting in cells with abnormal chromosome numbers.

Aneuploid

A cell or organism with an abnormal number of chromosomes due to the gain or loss of one or more chromosomes.

Monosomy

A type of aneuploidy in which one chromosome of a homologous pair is missing ($2n - 1$).

Trisomy

A type of aneuploidy in which an extra copy of a chromosome is present ($2n + 1$).

Gene Dosage

The number of copies of a particular gene present in a cell or organism, which influences the amount of gene product produced.

Down Syndrome (Trisomy 21)

A genetic disorder caused by the presence of an extra copy of chromosome 21.

X Inactivation

The process in female mammals in which one X chromosome is randomly condensed and transcriptionally silenced to equalize gene expression between males and females.

Barr Body

A condensed, inactive X chromosome found in the nuclei of most female mammalian cells.

Polyploidy

The presence of more than two complete sets of chromosomes in a cell or organism.

Chromosomal Rearrangement

A structural alteration of a chromosome resulting from chromosome breakage and rejoining.

Deletion

The loss of a segment of a chromosome, resulting in the removal of genetic material.

Duplication

A chromosomal mutation in which a segment of DNA is copied one or more additional times.

Inversion

A chromosomal rearrangement in which a chromosome segment breaks, rotates 180°, and reinserts into the chromosome.

Pericentric Inversion

An inversion that includes the centromere.

Paracentric Inversion

An inversion that does not include the centromere.

Translocation

A chromosomal rearrangement in which a chromosome segment moves to a different, nonhomologous chromosome.

Reciprocal Translocation

A translocation in which chromosome segments are exchanged between two nonhomologous chromosomes.

Check Your Understanding

1. What is nondisjunction, and during which stages of meiosis can it occur?
2. Compare nondisjunction occurring during Meiosis I with nondisjunction occurring during Meiosis II.
3. What is the difference between euploidy and aneuploidy?
4. Compare monosomy and trisomy.
5. Why are most autosomal monosomies incompatible with life?
6. What is X inactivation, and why is it necessary in female mammals?
7. What is a Barr body?
8. Explain why individuals with abnormal numbers of X chromosomes often have milder symptoms than individuals with autosomal aneuploidies.
9. Why are tortoiseshell cats almost always female?
10. What is polyploidy?
11. Why are triploid animals usually sterile?
12. Compare deletions and duplications.
13. How can changes in gene dosage affect phenotype?
14. What causes Cri-du-chat syndrome?
15. Compare pericentric and paracentric inversions.
16. Why do many inversions have relatively mild effects on phenotype?
17. What is a translocation?
18. Compare reciprocal and nonreciprocal translocations.
19. What is the Philadelphia chromosome, and why is it medically important?

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.

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [2.1.6: Errors in Meiosis](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [7.3: Errors in Meiosis](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

2.2: TRANSLATION

2.2.1: The Genetic Code

2.2.2: Ribosomes and Protein Synthesis

2.2.1: The Genetic Code

The process of transcription generates mRNA, a mobile copy of one or more genes with an alphabet of A, C, G, and U. *Translation* of the mRNA on ribosomes 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” (Figure 4.1). Different amino acids have different chemistries (such as acidic versus basic, or polar and nonpolar) and different structural constraints. Variation in amino acid sequence is responsible for the enormous variation in protein structure and function.

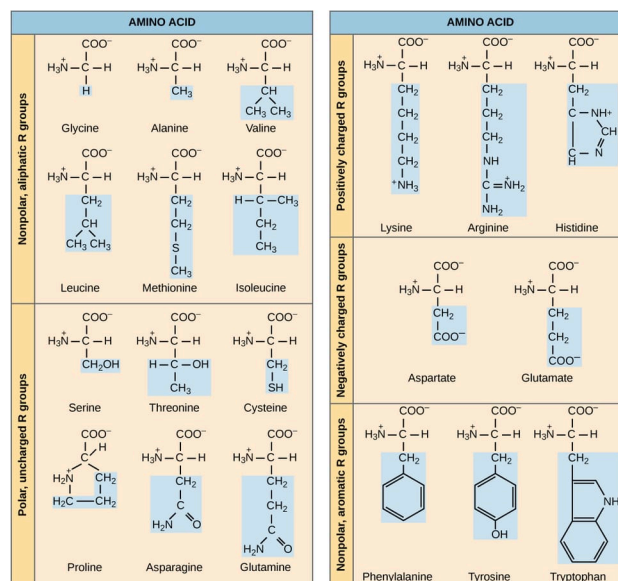


Figure 4.1 Structures of the 20 amino acids found in proteins are shown. Each amino acid is composed of an amino group (NH_3^+), a carboxyl group (COO^-), and a side chain (blue). The side chain may be nonpolar, polar, or charged, as well as large or small. It is the variety of amino acid side chains that gives rise to the incredible variation of protein structure and function.

The Central Dogma: RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the central dogma, as previously discussed. Genes specify the sequence of mRNAs, which in turn specify the sequence of amino acids making up all proteins. The decoding of one molecule to another is performed by specific proteins and RNAs. Because the information stored in DNA is so central to cellular function, it makes sense that the cell would make mRNA copies of this information for protein synthesis, while keeping the DNA itself intact and protected. The copying of DNA to RNA is relatively straightforward, as we saw in the previous chapter, with one nucleotide being added to the mRNA strand for every nucleotide read in the DNA strand. The *translation* to protein is a bit more complex because three mRNA nucleotides correspond to one amino acid in the polypeptide sequence. However, the translation to protein is still systematic and colinear, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.

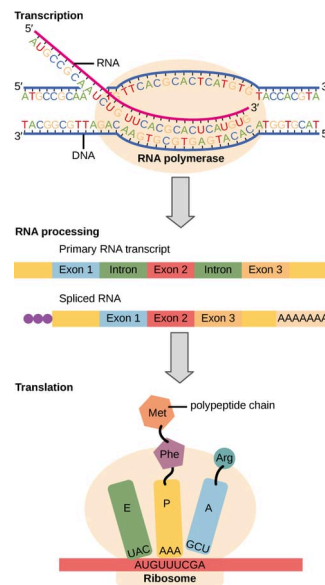


Figure 4.2 Instructions on DNA are transcribed onto messenger RNA. Ribosomes are able to read the genetic information inscribed on a strand of messenger RNA and use this information to string amino acids together into a protein.

The Genetic Code Is Degenerate and Universal

Because the mRNA 'alphabet' consists of only four nucleotide bases while the protein 'alphabet' consists of 20 amino acids, scientists realized that single letters could not specify individual amino acids. Combinations of two nucleotides (doublets) are also insufficient, yielding only 16 possible arrangements (4^2). In contrast, combinations of three nucleotides (triplets) yield 64 possible arrangements (4^3), providing far more than enough combinations to encode the 20 universal amino acids. This mathematical surplus led researchers to theorize that the genetic code must be **degenerate**, meaning a single amino acid can be specified by more than one distinct nucleotide triplet, now known as a **codon**.

This triplet hypothesis was famously confirmed by Francis Crick and Sydney Brenner using the chemical mutagen proflavin to insert nucleotides into a viral gene. Translation relies on a continuous **reading frame** composed of these triplets. When Crick and Brenner inserted one or two nucleotides, it shifted the entire downstream sequence, completely altering the message for every subsequent amino acid and preventing functional proteins from forming. However, when they inserted exactly three nucleotides, the reading frame was restored downstream; the resulting protein simply gained a single extra amino acid but otherwise maintained its structural integrity. This experiment definitively proved that the genetic code is read as a continuous series of *non-overlapping* nucleotide *triplets*. Scientists painstakingly solved the genetic code by translating synthetic mRNAs in vitro and sequencing the proteins they specified (Figure 4.4).

Frameshift Mutations

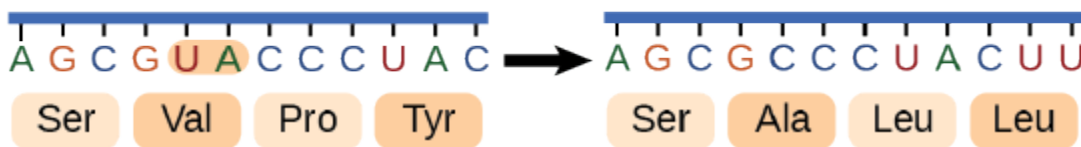


Figure 4.3 The deletion of two nucleotides shifts the reading frame of an mRNA and changes the entire protein message, creating a nonfunctional protein or terminating protein synthesis altogether.

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Figure 4.4 This figure shows the genetic code for translating each nucleotide triplet in mRNA into an amino acid or a termination signal in a protein. (credit: modification of work by NIH)

In addition to codons that instruct the addition of a specific amino acid to a polypeptide chain, three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **nonsense codons**, or **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. Following the start codon, the mRNA is read in groups of three until a stop codon is encountered.

The arrangement of the coding table reveals the structure of the code. There are sixteen "blocks" of codons, each specified by the first and second nucleotides of the codons within the block, e.g., the "AC*" block that corresponds to the amino acid threonine (Thr). Some blocks are divided into a pyrimidine half, in which the codon ends with U or C, and a purine half, in which the codon ends with A or G. Some amino acids get a whole block of four codons, like alanine (Ala), threonine (Thr) and proline (Pro). Some get the pyrimidine half of their block, like histidine (His) and asparagine (Asn). Others get the purine half of their block, like glutamate (Glu) and lysine (Lys). Note that some amino acids get a block and a half-block for a total of six codons.

Degeneracy is believed to be a cellular mechanism to reduce the negative impact of random mutations. Codons that specify the same amino acid typically only differ by one nucleotide. In addition, amino acids with chemically similar side chains are encoded by similar codons. For example, aspartate (Asp) and glutamate (Glu), which occupy the GA* block, are both negatively charged. This nuance of the genetic code ensures that a single-nucleotide substitution mutation might specify the same amino acid but have no effect or specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

The genetic code is nearly universal. With a few minor exceptions, virtually all species use the same genetic code for protein synthesis. Conservation of codons means that a purified mRNA encoding the globin protein in horses could be transferred to a tulip cell, and the tulip would synthesize horse globin. That there is only one genetic code is powerful evidence that all of life on Earth shares a common origin, especially considering that there are about 10^{84} possible combinations of 20 amino acids and 64 triplet codons.

The Wobble Hypothesis

The discovery that the genetic code is degenerate raises an important structural question. If there are 61 distinct codons that specify amino acids, does a cell require 61 unique tRNA molecules, each carrying a perfectly complementary anticodon? Experimentally, researchers discovered that most cells only possess between 30 to 45 distinct types of tRNAs. To explain how this smaller pool of tRNAs can accurately read all 61 codons, Francis Crick proposed the **Wobble Hypothesis** in 1966.

The hypothesis focuses on the geometry of codon-anticodon base pairing during translation. When an mRNA codon pairs with a tRNA anticodon, the first two nucleotide positions (reading in the 5' to 3' direction on the mRNA) must follow strict Watson-Crick base-pairing rules (A-U and G-C). However, the alignment at the third nucleotide position, the **3' position of the mRNA codon** and the matching **5' position of the tRNA anticodon**, is structurally flexible, or loose.

Because of this physical flexibility, non-standard hydrogen bonding can occur at this third position. For example, a guanine (G) at the wobble position of a tRNA anticodon can pair normally with a cytosine (C), but it can also flexibly pair with a uracil (U) on the mRNA. Furthermore, many tRNAs contain a modified nucleotide at their wobble position called **inosine (I)**, which can form stable hydrogen bonds with U, C, or A. This flexible "wobbling" allows a single tRNA molecule to recognize and translate multiple different mRNA codons, dramatically reducing the number of unique tRNA genes an organism needs to maintain in its genome while perfectly preserving the accuracy of protein synthesis.

Scientific Method Connection

Which Has More DNA: A Kiwi or a Strawberry?

Question: Would a kiwi and strawberry that are approximately the same size also have approximately the same amount of DNA?

Background: Genes are carried on chromosomes and are made of DNA. All mammals are diploid, meaning they have two copies of each chromosome. However, not all plants are diploid. The common strawberry is octoploid ($8n$) and the cultivated kiwi is hexaploid ($6n$). Research the total number of chromosomes in the cells of each of these fruits and think about how this might correspond to the amount of DNA in these fruits' cell nuclei. What other factors might contribute to the total amount of DNA in a single fruit? Read about the technique of DNA isolation to understand how each step in the isolation protocol helps liberate and precipitate DNA.

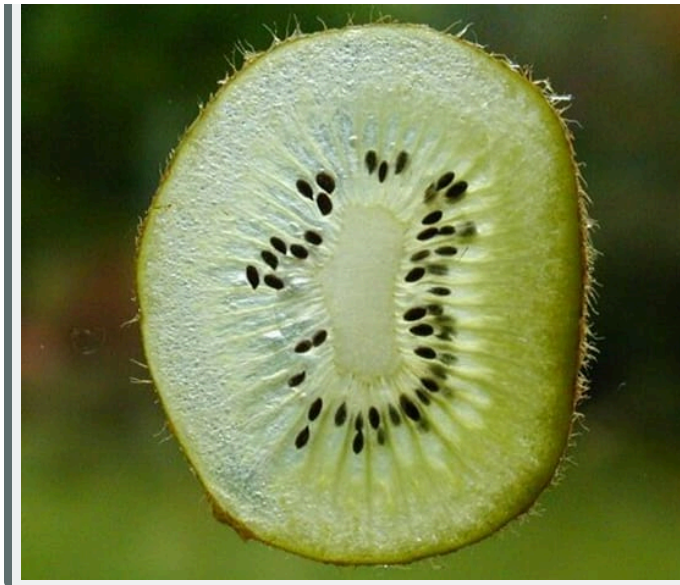
Hypothesis: Hypothesize whether you would be able to detect a difference in DNA quantity from similarly sized strawberries and kiwis. Which fruit do you think would yield more DNA?

Test your hypothesis: Isolate the DNA from a strawberry and a kiwi that are similarly sized. Perform the experiment in at least triplicate for each fruit

1. Prepare a bottle of DNA extraction buffer from 900 mL water, 50 mL dish detergent, and two teaspoons of table salt. Mix by inversion (cap it and turn it upside down a few times).
2. Grind a strawberry and a kiwi by hand in a plastic bag, or using a mortar and pestle, or with a metal bowl and the end of a blunt instrument. Grind for at least two minutes per fruit.
3. Add 10 mL of the DNA extraction buffer to each fruit, and mix well for at least one minute.
4. Remove cellular debris by filtering each fruit mixture through cheesecloth or porous cloth and into a funnel placed in a test tube or an appropriate container.
5. Pour ice-cold ethanol or isopropanol (rubbing alcohol) into the test tube. You should observe white, precipitated DNA.
6. Gather the DNA from each fruit by winding it around separate glass rods.

Analyze your data: Did you notice an obvious difference in the amount of DNA produced by each fruit? Were your results reproducible?

Draw a conclusion: Given what you know about the number of chromosomes in each fruit, can you conclude that chromosome number necessarily correlates to DNA amount? Can you identify any drawbacks to this procedure? If you had access to a laboratory, how could you standardize your comparison and make it more quantitative?



Key Terms

Genetic Code

The set of rules that specifies how nucleotide sequences are translated into amino acid sequences.

Codon

A sequence of three nucleotides in mRNA that specifies an amino acid or a translation signal.

Start Codon

A codon that signals the beginning of translation; usually AUG.

Stop Codon

A codon that signals the end of translation (UAA, UAG, or UGA).

Reading Frame

The grouping of nucleotides into consecutive codons during translation.

Degeneracy

The property that multiple codons can specify the same amino acid.

Universal Genetic Code

The observation that nearly all organisms use the same codon assignments.

Frameshift Mutation

An insertion or deletion of a nucleotide quantity not divisible by three, causing a disruption of the downstream reading frame.

Wobble Hypothesis

The principle that strict Watson-Crick base-pairing constraints are relaxed at the third position of a codon-anticodon, allowing a single tRNA to interact with multiple codons.

Sense Codon

Any codon that codes for an amino acid.

Nonsense Codon

A stop codon, signaling the end of translation.

Check Your Understanding

1. Why does the genetic code use three nucleotides per codon instead of one or two?
2. What is a codon?
3. What are the functions of the start codon?
4. Name the three stop codons.
5. What is a reading frame, and why is it important?
6. What does it mean that the genetic code is degenerate?
7. Why is the genetic code considered nearly universal?
8. Why are codons read from mRNA rather than DNA?

This page titled [2.2.1: The Genetic Code](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.2.2: Ribosomes and Protein Synthesis

The synthesis of proteins consumes more of a cell's energy than any other metabolic process. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform virtually every function of a cell. The process of **translation**, or protein synthesis, involves the decoding of an mRNA message into a polypeptide product. Amino acids are covalently strung together by interlinking peptide bonds in lengths ranging from approximately 50 to more than 1000 amino acid residues. Each individual amino acid has an amino group (NH₂) and a carboxyl (COOH) group (Figure 4.5). Polypeptides are formed when the amino group of one amino acid forms an amide (i.e., peptide) bond with the carboxyl group of another amino acid (Figure 4.5). This reaction is catalyzed by ribosomes and generates one water molecule.

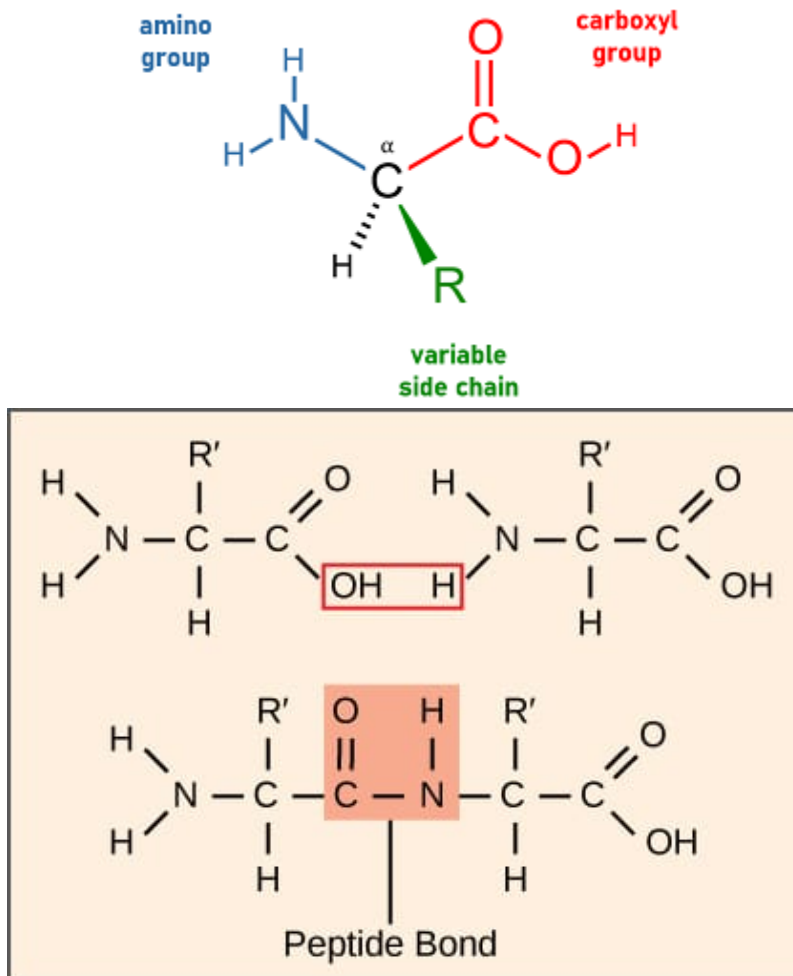


Figure 4.5 The general structure of an amino acid (top). A peptide bond links the carboxyl end of one amino acid with the amino end of another, producing one water molecule during the process (bottom). For simplicity in this image, only the functional groups involved in the peptide bond are shown. The R and R' designations refer to the rest of each amino acid structure.

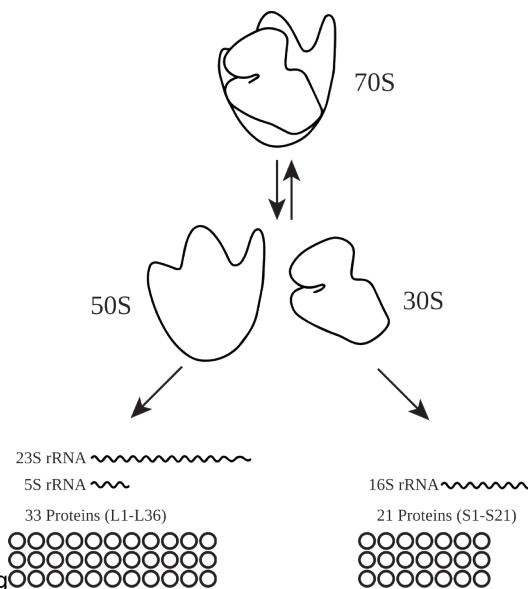
The Protein Synthesis Machinery

In addition to the mRNA template, many molecules and macromolecules contribute to the process of translation. The composition of each component may vary across species; for example, ribosomes may consist of different numbers of rRNAs 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. (Note: A ribosome can be thought of as an enzyme whose amino acid binding sites are specified by mRNA.)

Ribosomes

Before an mRNA is translated, a cell must invest energy to build each of its ribosomes. In *E. coli*, there are between 10,000 and 70,000 ribosomes present in each 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 exist in the cytoplasm of prokaryotes and in the cytoplasm and rough endoplasmic reticulum of eukaryotes. Mitochondria and chloroplasts also have their own ribosomes in the matrix and stroma, which look more similar to bacterial ribosomes (and have similar drug sensitivities) than the ribosomes just outside their outer membranes in the cytoplasm. Ribosomes dissociate into large and small subunits when they are not synthesizing proteins and reassociate during the initiation of translation. In *E. coli*, the small subunit is described as 30S, and the large subunit is 50S, for a total of 70S (recall that Svedberg units are not additive) (Figure 4.6). Mammalian ribosomes have a small 40S subunit and a large 60S subunit, for a total of 80S. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds tRNAs. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction: **reading the mRNA from 5' to 3' and synthesizing the polypeptide from the N terminus to the C terminus** (Figure 4.7). The complete mRNA/poly-ribosome structure is called a polysome.



File:70S_ribosome_composition.svg

Figure 4.6 Composition of the *Escherichia coli* 70S ribosome, including the rRNAs and proteins.

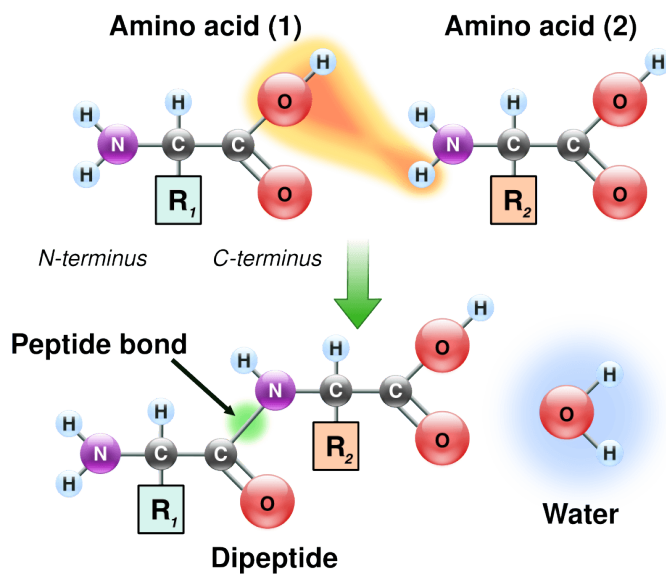


Figure 4.7. Peptide bond formation between two amino acids showing the N-terminus (amino group containing nitrogen) and the C-terminus (carboxyl group, containing carbon)

tRNAs

The tRNAs are structural RNA molecules that were transcribed from genes by RNA polymerase III. Depending on the species, 40 to 60 types of tRNAs exist in the cytoplasm. Transfer RNAs serve as *adaptor molecules*. Each tRNA carries a specific amino acid and recognizes one or more of the mRNA codons that define the order of amino acids in a protein. Aminoacyl-tRNAs bind to the ribosome 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, which is why they are called adaptors.

To achieve the high level of specificity, every tRNA molecule folds into a distinct three-dimensional inverted “L” shape (often flattened conceptually into a two-dimensional cloverleaf model) that features two primary functional domains at opposite ends of its structure. At the top of the molecule sits the 3' **acceptor stem**, which serves as the physical attachment site where a specific amino acid is chemically bound. At the exact opposite end is the **anticodon loop**, a specialized three-nucleotide sequence designed to directly read the genetic instructions written along the mRNA strand.

Of the 64 possible mRNA codons—or triplet combinations of A, U, G, and C—three specify the termination of protein synthesis and 61 specify the addition of amino acids to the polypeptide chain. Of these 61, one codon (**AUG**) also encodes the initiation of translation. Each tRNA anticodon can base pair with one or more of the mRNA codons for its amino acid. For instance, if the sequence 5' -CUA-3' occurred on an mRNA template in the proper reading frame, it would bind a leucine tRNA expressing the complementary, antiparallel anticodon sequence 3' -GAU-5'. This antiparallel alignment means that complementary nucleic acid strands must run in opposite physical directions, so the 5' end of the mRNA codon pairs directly with the 3' end of the tRNA anticodon.

As the adaptor molecules of translation, it is surprising that tRNAs can fit so much specificity into such a small package. Consider that tRNAs need to interact with three factors: 1) they must be recognized by the correct aminoacyl synthetase (see below); 2) they must be recognized by ribosomes; and 3) they must bind to the correct sequence in mRNA.

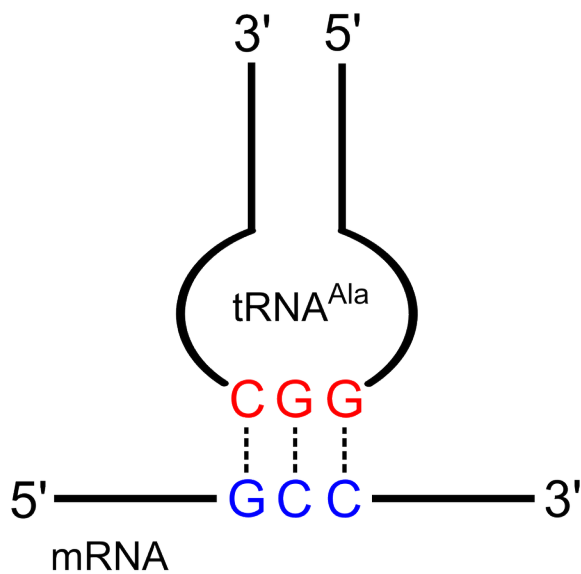


Figure 4.8 Codon-anticodon pairing exemplified for a tRNA^{Ala}

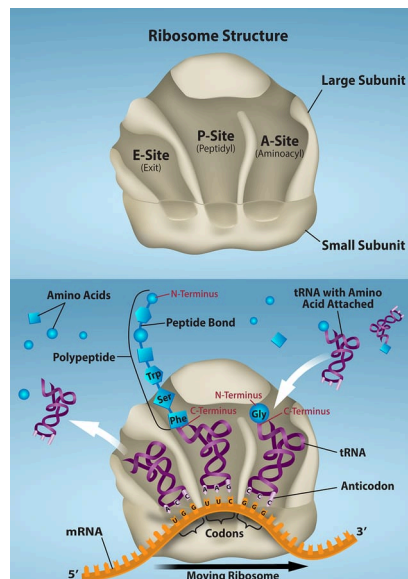


Figure 4.9 The ribosome and its function. The ribosome is responsible for translating the mRNA into protein. A. The ribosome consists of a large and small ribosomal subunit. Assembly of the subunits on the mRNA forms three tRNA binding sites. B. During translation, charged tRNAs enter the Acceptor site, and the anticodon on the tRNA base pairs with the codon in the mRNA. After the incoming amino acid forms a peptide bond with the growing polypeptide chain, the ribosome will move three nucleotides toward the 3' end of the mRNA. This movement will transfer the tRNA with the growing polypeptide to the Peptidyl-tRNA binding site and allow the empty tRNA to exit at the Exit site. Credit: Rao, A., Ryan, K. and Fletcher, S. Department of Biology, Texas A&M University.

Aminoacyl tRNA Synthetases

The process of pre-tRNA synthesis by RNA polymerase III only creates the RNA portion of the adaptor molecule. The corresponding amino acid must be added later, once the tRNA is processed and exported to the cytoplasm. Through the process of **tRNA “charging”**, each tRNA molecule is linked to its correct amino acid by one of a group of enzymes called **aminoacyl tRNA synthetases**. At least one type of aminoacyl tRNA synthetase exists for each of the 20 amino acids; the exact number of aminoacyl tRNA synthetases varies by species. These enzymes first bind and hydrolyze ATP to catalyze a high-energy bond between an amino acid and adenosine monophosphate (AMP); a pyrophosphate molecule is expelled in this reaction. The

activated amino acid is then transferred to the tRNA, and AMP is released. The term "charging" is appropriate, since the high-energy bond that attaches an amino acid to its tRNA is later used to drive the formation of the peptide bond. Each tRNA is named for its amino acid.

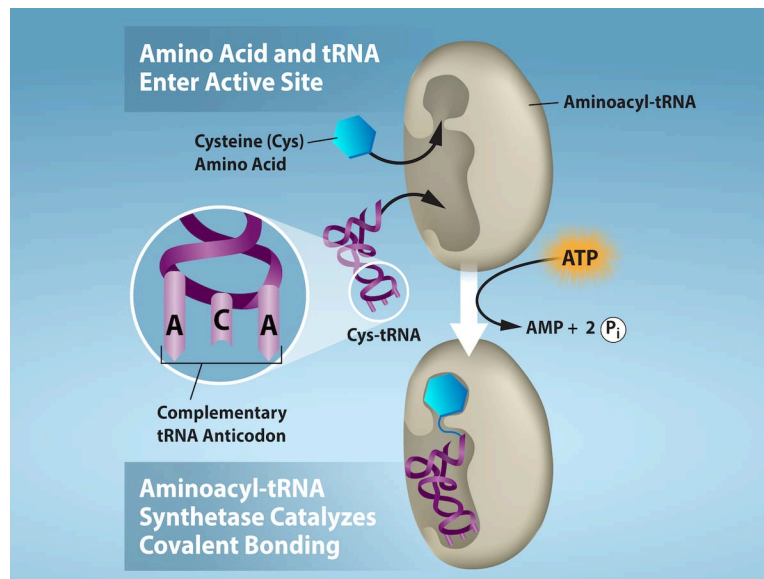


Figure 4.10 Charging of tRNAs with correct amino acids. Aminoacyl-tRNA synthetases catalyze covalent bond formation between the tRNA and the correct amino acid in preparation for translation. Because there are multiple amino acids, there are multiple different tRNA synthetases. All of the synthetases require energy, in the form of ATP, to make sure the correct amino acid is attached to the tRNA with the correct anticodon sequence. Credit: Rao, A., Ryan, K. and Tag, A. Department of Biology, Texas A&M University.

The Mechanism of Protein Synthesis

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'll explore how translation occurs in *E. coli*, and specify any differences between bacterial and eukaryotic translation.

Initiation of Translation

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves the small 30S ribosomal subunit, the mRNA template, three initiation factors (IFs; IF-1, IF-2, and IF-3), and a special initiator tRNA, called **tRNA^{Metf}**.

In *E. coli* mRNA, a sequence upstream of the first AUG codon, called the **Shine-Dalgarno sequence (AGGAGG)**, interacts with the 16 S rRNA molecule of the small ribosomal subunit. This interaction anchors the 30S ribosomal subunit at the correct location on the mRNA template. Guanosine triphosphate (GTP), which is a purine nucleotide triphosphate, acts as an energy source during translation—both at the start of elongation and during the ribosome's translocation. Binding of the mRNA to the 30S ribosome also requires IF-III.

The initiator tRNA^{Metf} then interacts with the start codon AUG (or rarely, GUG). This tRNA carries the amino acid methionine, which is formylated after its attachment to the tRNA. The formylation creates a "faux" peptide bond between the formyl carboxyl group and the amino group of the methionine. Binding of the fMet-tRNA^{Metf} is mediated by the initiation factor IF-2. The fMet begins every polypeptide chain synthesized by *E. coli*, but it is usually removed after translation is complete. When an in-frame AUG is encountered during translation elongation, a non-formylated methionine is inserted by a regular Met-tRNA^{Met}. After the formation of the initiation complex, the 30S ribosomal subunit is joined by the 50S subunit to form the translation complex. In eukaryotes, a similar initiation complex forms, comprising mRNA, the 40S small ribosomal subunit, eukaryotic IFs, and nucleoside triphosphates (GTP and ATP). The methionine on the charged initiator tRNA, called Met-tRNA_i, is not formylated. However, Met-tRNA_i is distinct from other Met-tRNAs in that it can bind IFs.

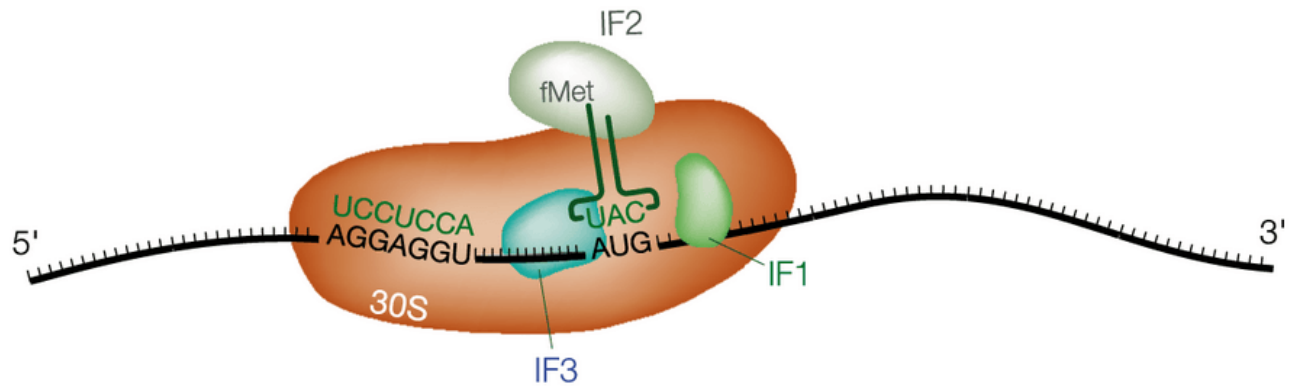


Figure 4.11 Initiation of protein translation in bacteria. The 30S binds to the mRNA through interaction between the purine-rich Shine-Dalgarno sequence and the 3'-end of the 16rRNA. IF1, IF2 and IF3 participate in proper initiation codon recognition.

Eukaryotes do not have the Shine-Dalgarno sequence, instead the eukaryotic initiation complex recognizes the 7-methylguanosine cap at the 5' end of the mRNA. A cap-binding protein (CBP) and several other IFs assist the movement of the ribosome to the 5' cap. Once at the cap, the initiation complex tracks along the mRNA in the 5' to 3' direction, searching for the AUG start codon. Many eukaryotic mRNAs are translated from the first AUG, but this is not always the case. According to Kozak's rules, the nucleotides around the AUG indicate whether it is the correct start codon. Kozak's rules state that the following consensus sequence must appear around the AUG of vertebrate genes: 5'-gccRccAUGG-3'. This is also called the **Kozak sequence**. The R (for purine) indicates a site that can be either A or G, but cannot be C or U. Essentially, the closer the sequence is to this consensus, the higher the efficiency of translation.

Once the appropriate AUG is identified, the other proteins and CBP dissociate, and the 60S subunit binds to the complex of Met-tRNA_i, mRNA, and the 40S subunit. This step completes the initiation of translation in eukaryotes.

Translation, Elongation, and Termination

In prokaryotes and eukaryotes, the basics of elongation are the same, so we will review elongation from the perspective of *E. coli*. When the translation complex is formed, the tRNA binding region of the ribosome consists of three compartments (refer to figure 4.9). The **A (aminoacyl) site** binds incoming charged aminoacyl tRNAs. The **P (peptidyl) site** binds charged tRNAs carrying amino acids that have formed peptide bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The **E (exit) site** releases dissociated tRNAs so that they can be recharged with free amino acids. The initiating methionyl-tRNA, however, occupies the P site at the beginning of the elongation phase of translation in both prokaryotes and eukaryotes.

During translation elongation, the mRNA template provides tRNA binding specificity. As the ribosome moves along the mRNA, each mRNA codon comes into register, and specific binding with the corresponding charged tRNA anticodon is ensured. If mRNA were not present in the elongation complex, the ribosome would bind tRNAs nonspecifically and randomly.

Elongation proceeds with charged tRNAs sequentially entering and leaving the ribosome as each new amino acid is added to the polypeptide chain. Movement of a tRNA from A to P to E site is induced by conformational changes that advance the ribosome by three bases in the 3' direction. The energy for each step along the ribosome is donated by elongation factors that hydrolyze GTP. GTP energy is required both for the binding of a new aminoacyl-tRNA to the A site and for its translocation to the P site after formation of the peptide bond. Peptide bonds form between the amino group of the amino acid attached to the A-site tRNA and the carboxyl group of the amino acid attached to the P-site tRNA. The formation of each peptide bond is catalyzed by **peptidyl transferase**, an RNA-based enzyme that is integrated into the 50S ribosomal subunit. The energy for each peptide bond formation is derived from the high-energy bond linking each amino acid to its tRNA. *After peptide bond formation, the A-site tRNA that now holds the growing peptide chain moves to the P site, and the P-site tRNA that is now empty moves to the E site and is expelled from the ribosome (Figure 15.18).* Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino-acid protein can be translated in just 10 seconds.

Visual Connection

5'-CUUAAAUGCUCAAAGUAAACUGUGUCUAAUU-3'

Figure 4.12 Translation begins when an initiator tRNA anticodon recognizes a start codon on mRNA bound to a small ribosomal subunit. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, successive tRNAs move through the ribosome and the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Many antibiotics inhibit bacterial protein synthesis. For example, tetracycline blocks the A site on the bacterial ribosome, and chloramphenicol blocks peptidyl transferase. What specific effect would you expect each of these antibiotics to have on protein synthesis?

Tetracycline would directly affect:

- tRNA binding to the ribosome
- ribosome assembly
- growth of the protein chain

Chloramphenicol would directly affect:

- tRNA binding to the ribosome
- ribosome assembly
- growth of the protein chain

Termination of translation occurs when a nonsense codon (UAA, UAG, or UGA) is encountered. Upon aligning with the A site, these nonsense codons are recognized by **protein release factors** that resemble tRNAs. The releasing factors in both prokaryotes and eukaryotes instruct peptidyl transferase to add a water molecule to the carboxyl end of the P-site amino acid. This reaction forces the P-site amino acid to detach from its tRNA, and the newly made protein is released. The small and large ribosomal subunits dissociate from the mRNA and from each other; they are recruited almost immediately into another translation initiation complex. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

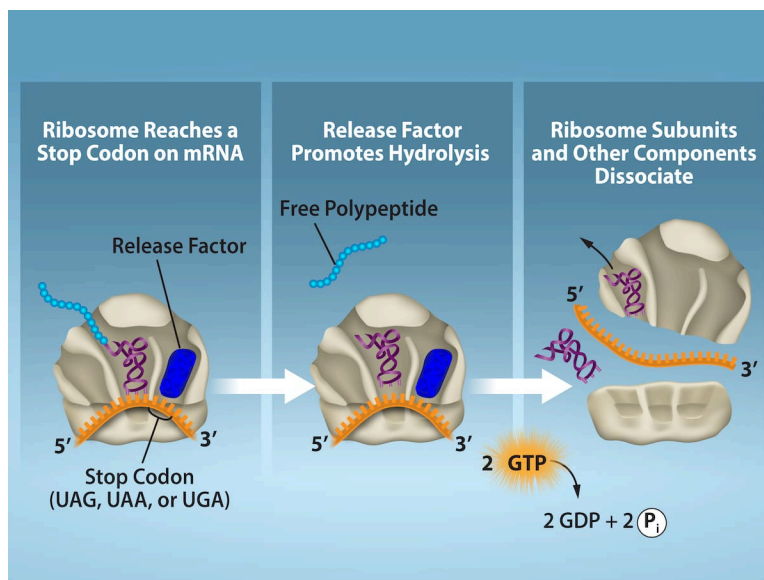


Figure 4.13 Translation Termination is Active. Translation is terminated when a STOP codon is in the A-site of the ribosome. Since there are no tRNAs corresponding to the STOP codons, the Release Factor protein enters and catalyzes the hydrolysis between the last amino acid and its tRNA. This hydrolysis releases the free carboxyl terminus (C-term) of the protein. Additional factors use the energy in GTP hydrolysis to disassemble the large and small ribosomal subunits and mRNA. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

Protein Folding, Modification, and Targeting

During and after translation, individual amino acids may be chemically modified, signal sequences appended, and the new protein “folded” into a distinct three-dimensional tertiary structure as a result of intramolecular interactions. These chemical alterations, known as post-translational modifications, fundamentally change a protein's behavior, stability, and function after its amino acid chain has been successfully assembled. For instance, phosphorylation involves the chemical addition of a phosphate group to specific amino acids, acting as a molecular on/off switch to activate or deactivate the protein's functional state. Similarly, glycosylation features the covalent attachment of carbohydrate chains to convert the polypeptide into a glycoprotein, a modification that is essential for cell membrane structural integrity and cellular recognition. In other instances, proteolytic cleavage intentionally cuts the peptide backbone via specialized enzymes as a maturation step that transforms a precursor molecule into an active protein or hormone, such as the cleavage of proinsulin into functional insulin.

A **signal sequence** may be added, which is a short sequence at the amino end of a protein that directs it to a specific cellular compartment. These sequences can be thought of as the protein's “train ticket” to its ultimate destination, and they are recognized by signal-recognition proteins (SRP) that act as conductors. Proteins destined for the cytoplasm, mitochondria, or chloroplasts in plants are fully translated directly within the cytosol before their specific signal sequence terminus guides them to their target organelle. Conversely, proteins bound for the cell membrane, lysosomes, or eventual secretion outside the cell are targeted co-translationally to the rough endoplasmic reticulum (RER). The SRP binds the nascent signal sequence during translation, temporarily pausing the assembly process, and moves the entire ribosome complex to dock directly onto the RER membrane. The growing protein is then threaded straight inside the RER lumen, modified, and packaged into transport vesicles bound for the Golgi apparatus, which operates as the cell's final sorting and shipping hub where proteins are clipped of their temporary signal sequences upon arrival.

While many proteins fold spontaneously into their native states, others need helper molecules called **chaperones** to prevent them from mis-folding or aggregating during this transition. Even if a protein sequence is properly specified by its corresponding mRNA, it can easily take on a completely dysfunctional, denatured shape if abnormal temperature or pH conditions disrupt the intramolecular interactions required for correct folding. When a protein permanently mis-folds due to environmental or cellular stress, the cell activates a quality control pathway to prevent toxic intracellular aggregation. Mis-folded or damaged proteins are systematically tagged with a small regulatory protein called **ubiquitin**. This molecular “trash tag” targets the defective polypeptide for immediate destruction inside the proteasome, a barrel-shaped cellular complex that shreds the broken chain back into raw amino acids, cleanly recycling them for future rounds of translation.

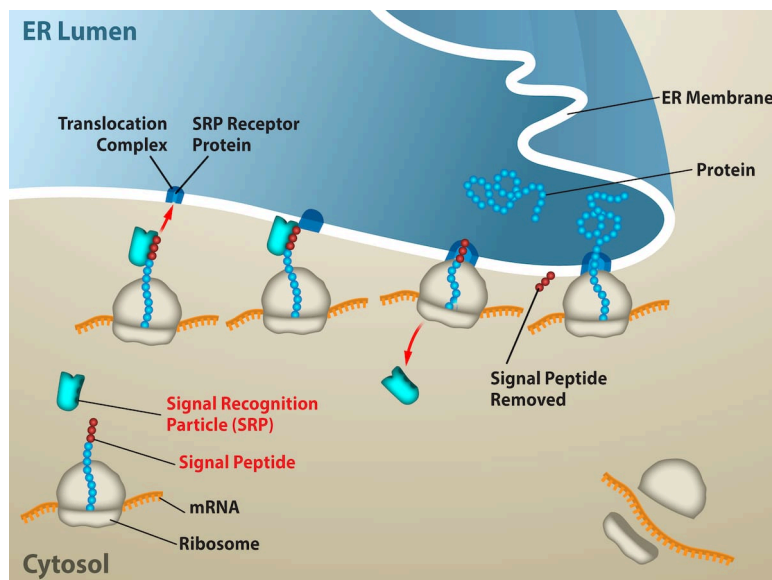


Figure 4.14 Proteins are co-translationally targeted into the ER for secretion. Proteins that will be secreted from the cell will contain a signal sequence at the N-terminus. The signal will be recognized by SRP as soon as the amino acids emerge from the ribosome, and the ribosome will be targeted to the translocation channel in the ER membrane. The rest of the protein will go directly from the ribosome, across the ER membrane and into the ER lumen. From the ER, proteins can be secreted from the cell via vesicle trafficking. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

Translation in Archaea

The ribosomes of Archaea closely resemble bacteria. They possess 70S ribosomes composed of 30S and 50S subunits, and rely heavily on a sequence highly similar to the bacterial Shine-Dalgarno sequence to align the ribosome at the proper start site. However, once the translation actually begins, the machinery shifts entirely toward the eukaryotic mechanism. Archaeal ribosomes contain a suite of unique structural proteins shared exclusively with eukaryotic ribosomes and completely absent in bacteria. Furthermore, while bacteria initiate translation using a specialized formyl-methionine amino acid, Archaea use unmodified methionine, just like eukaryotes. Archaea also completely lack standard bacterial initiation factors. Instead, they use an array of regulatory proteins that are structurally and functionally homologous to eukaryotic initiation factors.

Key Terms

Transfer RNA (tRNA)

An RNA molecule that carries an amino acid and recognizes a codon during translation.

Anticodon

A three-nucleotide sequence in tRNA that base-pairs with a complementary mRNA codon.

Aminoacyl-tRNA Synthetase

An enzyme that attaches the correct amino acid to the appropriate tRNA.

tRNA Charging

The process of attaching an amino acid to its corresponding tRNA.

Charged tRNA (Aminoacyl-tRNA)

A tRNA with an amino acid attached.

Shine-Dalgarno Sequence

A purine-rich leader sequence (5' -AGGAGG-3') on bacterial mRNA that guides the 30S ribosome to the correct start site.

Aminoacyl (A) Site

The first binding pocket in the ribosome, which accommodates incoming charged aminoacyl-tRNAs.

Peptidyl (P) Site

The central binding pocket in the ribosome that anchors the tRNA holding the growing polypeptide chain.

Exit (E) Site

The final pocket in the ribosome that releases empty, uncharged tRNAs after their amino acid has been donated..

Signal Sequence

A short stretch of amino acids at the amino end of a nascent protein that directs its transport to specific cellular destinations (e.g., the Rough ER).

Check Your Understanding

1. Why is tRNA described as a molecular adaptor?
2. What are the two major functional regions of a tRNA molecule?
3. What is the difference between a codon and an anticodon?
4. Why are codons and anticodons described as antiparallel?
5. Where is an amino acid attached to a tRNA?
6. What is the function of an aminoacyl-tRNA synthetase?
7. Why is accurate tRNA charging essential for protein synthesis?
8. Why are stop codons not recognized by amino acid-carrying tRNAs?
9. Which direction is an mRNA strand read by a ribosome, and in what direction is the resulting polypeptide synthesized?
10. During the elongation stage of translation, what are the three distinct tRNA binding sites on the large ribosomal subunit, and what occurs at each site?
11. How many distinct aminoacyl-tRNA synthetase enzymes must exist at a bare minimum in any given species to ensure proper translation?

2.3: GENE EXPRESSION

2.3.1: Regulation of Gene Expression

2.3.2: Prokaryotic Gene Regulation

2.3.3: Eukaryotic Gene Regulation

2.3.4: Cancer and Gene Regulation

2.3.1: Regulation of Gene Expression

For a cell to function properly, necessary proteins must be synthesized at the proper time and place. All cells control or regulate the synthesis of proteins. The process of "turning on" a gene to produce RNA and protein is called **gene expression**. Whether in a unicellular organism or a complex multi-cellular organism, each cell controls when and how its genes are expressed. For this to occur, there must be internal chemical mechanisms that control when a gene is expressed, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

The regulation of gene expression conserves energy and space. It would require a significant amount of energy for an organism to express every gene at all times, so it is more energy efficient to turn on the genes only when they are required. In addition, only expressing a subset of genes in each cell saves space because DNA must be unwound from its tightly coiled structure to transcribe and translate the DNA. Cells would have to be enormous if every protein were expressed in every cell all the time. 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 codes for a functional protein in a cell. The process occurs in both prokaryotic and eukaryotic cells, just in slightly different manners. Prokaryotic organisms lack a cell nucleus and their DNA therefore floats freely in the cell cytoplasm. To synthesize a protein, the processes of transcription and translation occur almost simultaneously. When the resulting protein is no longer needed, transcription stops. As a result, the primary method to control what type of protein and how much of each protein is expressed in a prokaryotic cell is the *regulation of transcription*. All of the subsequent steps occur automatically. When more protein is required, more transcription occurs. Therefore, in prokaryotic cells, the control of gene expression is mostly at the *transcriptional level*.

Eukaryotic cells, in contrast, have intracellular organelles that add to their complexity. In eukaryotic cells, the DNA is contained inside the cell's nucleus and there it is transcribed into RNA. The newly synthesized RNA is then transported out of the nucleus into the cytoplasm, where ribosomes translate the RNA into protein. The processes of transcription and translation are *physically separated* by the nuclear membrane; transcription occurs only within the nucleus, and translation occurs only outside the nucleus in the cytoplasm. The regulation of gene expression can occur at all stages of the process. 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 the 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**).

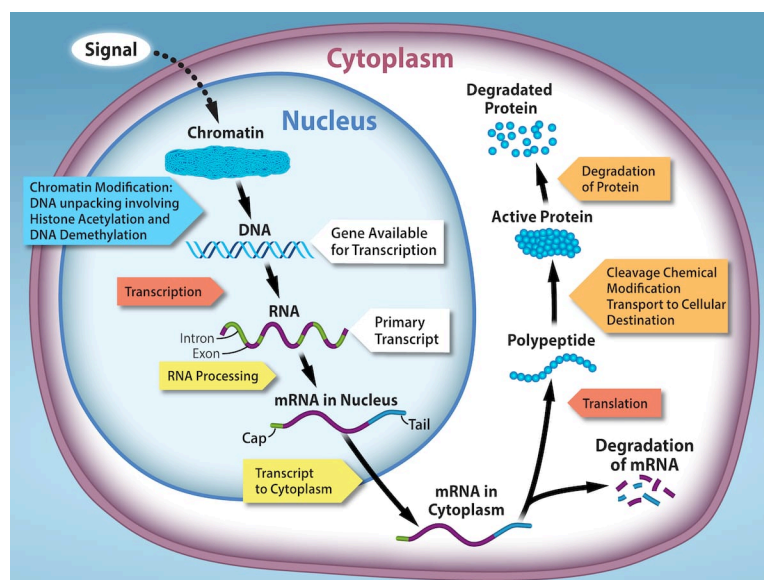


Figure 5.1 Locations of gene regulation. The regulation of gene expression occurs at multiple steps going from DNA to the functional gene product, usually a protein. It begins with chromatin structure making the DNA more or less accessible for transcription by RNA polymerase. In eukaryotes, the primary mRNA transcript must be processed before it can be translated in the cytoplasm. The final level of active protein in the cell depends not only on the rate of synthesis, but also on the rate of degradation of mRNA and protein. Credit: Rao, A. and Ryan, K. Department of Biology, Texas A&M University.

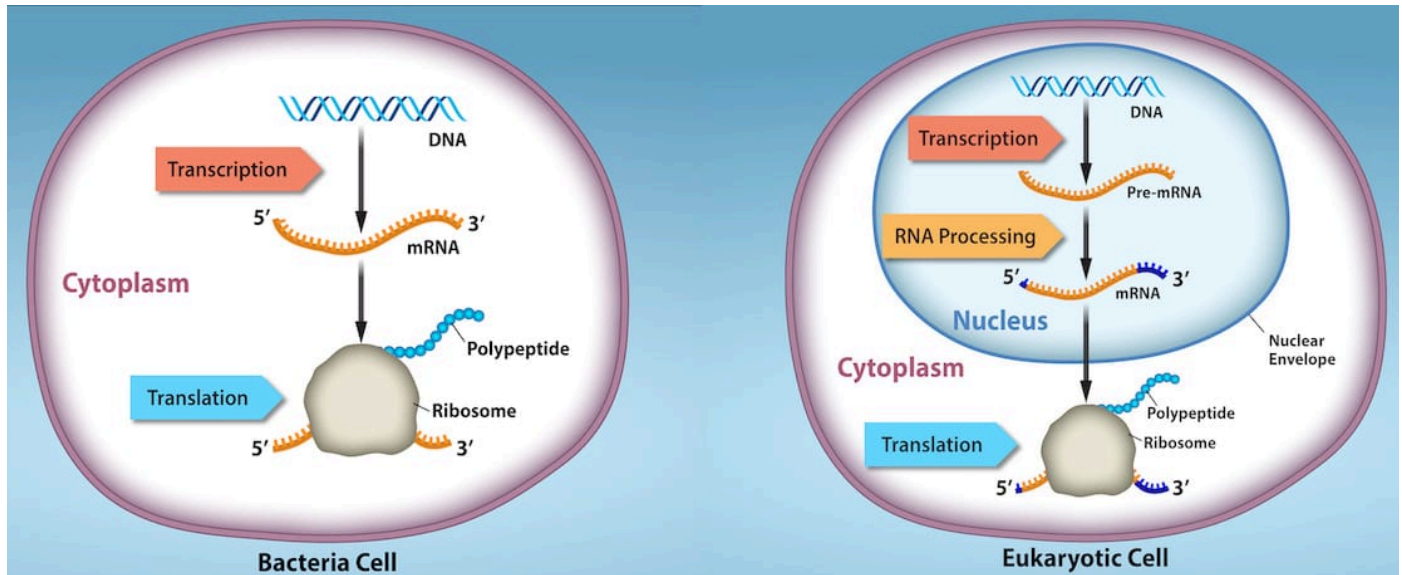


Figure 5.2 Regulation in prokaryotes and eukaryotes. A. Prokaryotic transcription and translation occur simultaneously in the cytoplasm, and regulation occurs primarily at the transcriptional level. B. Eukaryotic gene expression is regulated during transcription and RNA processing, which take place in the nucleus, and during protein translation, which takes place in the cytoplasm. Further regulation may occur through post-translational modifications of proteins in both prokaryotes and eukaryotes. Credit: Rao, A., Ryan, K. Fletcher, S. and Tag, A. Department of Biology, Texas A&M University.

2.3.2: Prokaryotic Gene Regulation

The DNA of prokaryotes is organized into a circular chromosome, supercoiled within the nucleoid region of the cell cytoplasm. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are encoded together in blocks called **operons**. An operon is a functional genetic unit composed of a promoter, an operator, and a continuous cluster of structural genes. The promoter is the upstream sequence where RNA polymerase physically binds to initiate transcription, while the **operator** is a downstream regulatory DNA sequence that acts as a gatekeeper. By organizing genes this way, all of the necessary enzymes for a specific pathway, such as the genes needed to use lactose as an energy source in the **lactose (lac) operon**, are arranged right next to each other and transcribed into a single, continuous polycistronic mRNA molecule.

To control when operons are active, cells use three types of regulatory molecules that affect expression: repressors, activators, and inducers. Repressors and activators are specialized regulatory proteins encoded by separate regulatory genes, and they control expression by physically binding to the specific DNA sites adjacent to the structural genes. **Activators** bind to or near the promoter site to recruit RNA polymerase, effectively increasing transcription in response to external stimuli. Conversely, **repressors** bind directly to the operator region. When a repressor locks onto the operator, it acts as a physical roadblock that prevents RNA polymerase from moving downstream along the DNA strand, halting transcription. The activity of these structural proteins is modulated by **inducers**, which are small effector molecules present in the cell or its environment. Inducers bind directly to the regulatory proteins, altering their physical shape to either trigger or stop their binding to the DNA.

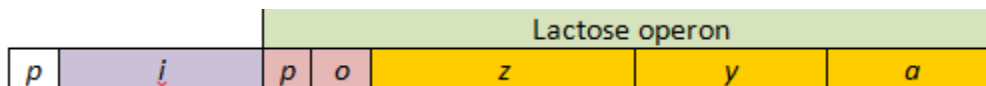


Figure 5.3 The lactose operon. The far left (P) is the promoter for the repressor gene (i). The Lac operon consists of the operon promoter (pink p), the operator (o), and the three structural genes: lac z, lac y, and lac a. The repressor gene codes for the repressor protein that sits on the operator preventing transcription until it is removed.

The *trp* Operon: A Repressible Operon

Bacteria such as *Escherichia coli* need a constant supply of amino acids to survive, and are able to synthesize many of them when they are scarce in the environment. Tryptophan is one such amino acid that *E. coli* can either ingest from the environment or synthesize using enzymes that are encoded by five genes. These five genes are next to each other in what is called the **tryptophan (trp) operon** (Figure 16.4). The genes are transcribed into a single mRNA, which is then translated to produce all five enzymes. If tryptophan is present in the environment, then *E. coli* does not need to synthesize it and the *trp* operon is switched off. However, when tryptophan availability is low, the switch controlling the operon is turned on, the mRNA is transcribed, the enzyme proteins are translated, and tryptophan is synthesized.

The *trp* operon is divided into three key regions: the upstream promoter sequence, the operator sequence, and the downstream protein-coding region. Positioned immediately between the promoter and the transcription start site, the operator sequence acts as the physical binding site for the *trp* repressor protein. Because it is an anabolic pathway designed to build a molecule, the *trp* operon operates as a **negative repressible system**, meaning it defaults to an "on" state and is actively switched "off" only when its end product accumulates. The word "negative" here refers strictly to the type of regulatory protein involved, meaning that transcription is controlled by a *repressor* that turns the system off, rather than an activator that turns it on. The word "repressible" refers to the operon's default state, meaning the system defaults to an "on" position and must be actively pushed "off" or *repressed* by an external stimulus, in this case, when its metabolic end product accumulates.

The repressor protein is synthesized in an *inactive conformation* and cannot bind to the operator on its own. However, when tryptophan levels rise within the cell, two tryptophan molecules physically bind to the repressor protein as **co-repressors**. This binding induces an allosteric conformational shift, altering the repressor's shape into an active state that tightly clamps onto the *trp* operator. This physical roadblock prevents RNA polymerase from transcribing the downstream structural genes. When intracellular tryptophan availability drops, the co-repressor molecules dissociate, the repressor reverts to its inactive shape and falls off the DNA, allowing RNA polymerase to resume transcription and replenish the cell's amino acid supply.

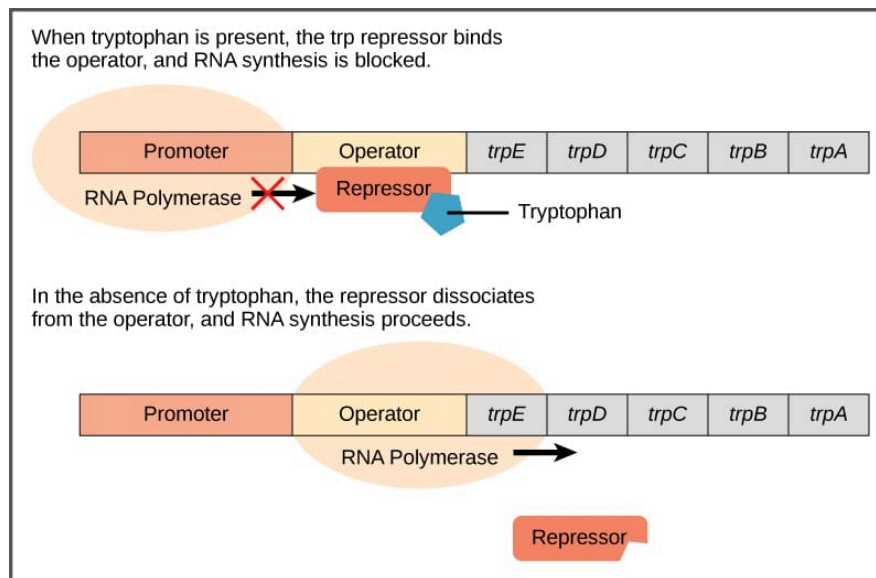


Figure 5.4 The tryptophan operon. The five genes that are needed to synthesize tryptophan in *E. coli* are located next to each other in the *trp* operon. When tryptophan is plentiful, two tryptophan molecules bind the repressor protein at the operator sequence. This physically blocks the RNA polymerase from transcribing the tryptophan genes. When tryptophan is absent, the repressor protein does not bind to the operator and the genes are transcribed.

Link to Learning

Watch this video to learn more about the *trp* operon.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



The *lac* Operon: An Inducible Operon

While repressible operons control anabolic pathways to stop production when a nutrient is abundant, **inducible** operons typically regulate catabolic pathways, remaining turned "off" until a specific substrate arrives to be broken down. The classic example for this mechanism is the lactose (*lac*) operon in *E. coli*, which encodes the enzymes required to acquire and digest lactose from the environment. The operon features three structural genes: *lacZ*, which encodes β -galactosidase to cleave lactose into glucose and galactose; *lacY*, which encodes lactose permease to pump the sugar into the cell; and *lacA*, which encodes a transacetylase enzyme.

Because synthesizing these enzymes is energetically costly, the *lac* operon defaults to an "off" state via a **negative inducible** control mechanism. In the absence of lactose, a constitutively expressed *lac* repressor protein binds tightly to the operator, physically blocking RNA polymerase from moving forward. However, when lactose becomes available, a small amount is converted into its metabolic isomer, *allolactose*. Allolactose acts as an *inducer*, binding to the *lac* repressor and forcing an allosteric structural change that causes the protein to release from the operator sequence, clearing the way for transcription.

Even when the repressor is removed, the cell faces a secondary energetic dilemma because glucose remains its preferred, most efficient fuel source. To prevent the wasteful digestion of lactose when glucose is plentiful, the *lac* operon is also controlled by a secondary, *positive* regulatory mechanism, the **Catabolite Activator Protein (CAP)**. When environmental glucose is

abundant, intracellular signaling molecules called cyclic AMP (cAMP) remain low. However, when glucose levels drop, cAMP rapidly accumulates inside the cytoplasm and binds directly to CAP, activating it.

Once activated by cAMP, the CAP complex binds to a specialized consensus site located immediately upstream of the *lac* promoter. The physical presence of CAP stabilizes the docking of RNA polymerase, increasing its binding affinity to an otherwise weak promoter sequence and boosting transcription rates exponentially. Full activation of the *lac* operon requires 2 things to happen: glucose must be completely absent (triggering cAMP/CAP activation to accelerate the engine) and lactose must be present (triggering allolactose induction to remove the physical brake). If glucose is present, the operon is restricted to a negligible, basal rate of background transcription, protecting the cell from wasting energy.

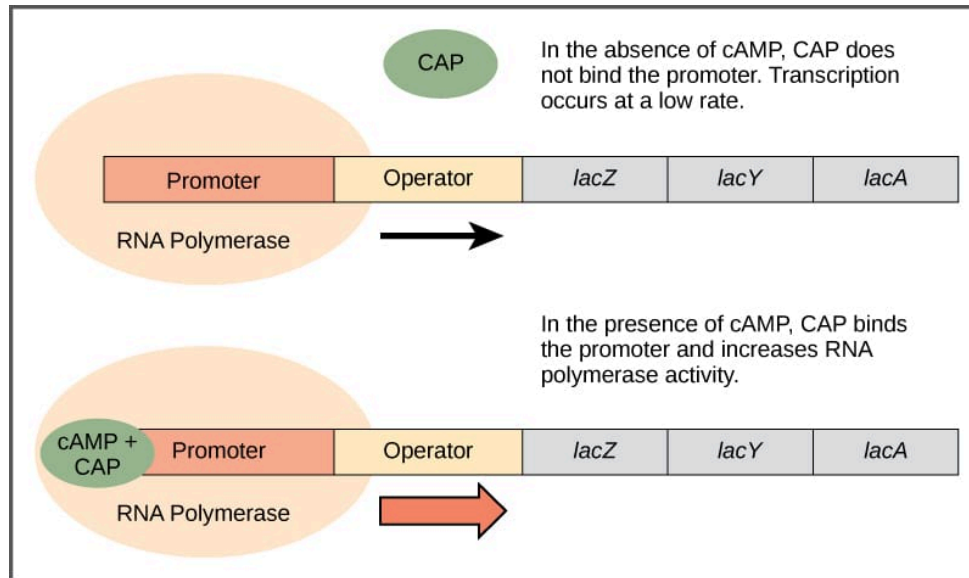


Figure 5.5 Transcriptional activation by the CAP protein. When glucose levels fall, *E. coli* may use lactose for fuel but must transcribe new genes to do so. As glucose supplies become limited, cAMP levels increase. This cAMP binds to the CAP protein, a positive regulator that binds to a promoter region upstream of the genes required to use lactose.

The *lac* operon serves as a textbook example of a system that is both **negative inducible** and **positive inducible**. The system is "negative" because its primary control switch is a *repressor protein* that blocks transcription, and it is "inducible" because its default metabolic state is "off" and must be actively turned "on" by an environmental inducer (allolactose). The system is also "positive" because it relies on an *activator protein* (CAP) to increase transcription to high levels, and it is "inducible" because this activator remains dormant until it is turned "on" by a signaling molecule (cAMP) indicating that glucose is absent.

Visual Connection

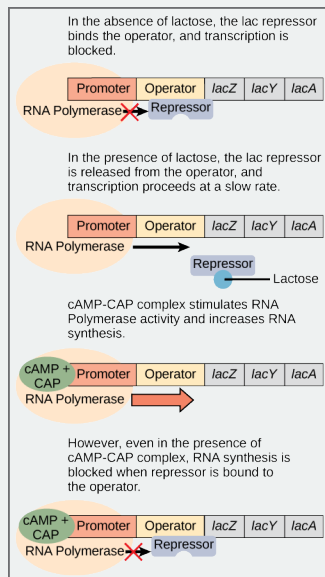


Figure 5.6 Regulation of the *lac* operon. Transcription of the *lac* operon is carefully regulated so that its expression only occurs when glucose is limited and lactose is present to serve as an alternative fuel source.

Signals that Induce or Repress Transcription of the *lac* Operon

Glucose	CAP binds	Lactose	Repressor binds	Transcription
+	-	-	+	No
+	-	+	-	Some
-	+	-	+	No
-	+	+	-	Yes

Table 5.1

Link to Learning

Watch an animated tutorial about the workings of *lac* operon here.

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Key Terms

Operon

Genetic unit found in prokaryotes composed of a promoter, an operator, and a continuous cluster of functionally related structural genes transcribed together into a single mRNA molecule.

Operator

A downstream regulatory DNA sequence situated between the promoter and the structural genes that acts as a physical

gatekeeper by binding repressor proteins.

Negative Regulation

A gene control mechanism in which transcription is governed by a *repressor protein* that physically blocks RNA polymerase from moving down the DNA strand.

Positive Regulation

A gene control mechanism in which transcription is governed by an *activator protein* that physically binds to the DNA to recruit and stabilize RNA polymerase at a promoter.

Repressible Operon

An operon whose default metabolic state is "on", which must be actively switched "off" or repressed when an environmental stimulus or metabolic end product accumulates (e.g., the trp operon).

Inducible Operon

An operon whose default metabolic state is "off", which must be actively switched "on" or induced when a specific environmental substrate arrives to be processed (e.g., the lac operon).

Catabolite Activator Protein (CAP)

A positive regulatory activator protein that, when bound to cyclic AMP (cAMP), locks onto a consensus site upstream of a promoter to exponentially accelerate transcription rates.

Check Your Understanding

1. In a standard prokaryotic operon, what is the linear sequence of the three primary regions from upstream to downstream?
2. How does a repressor protein stop a gene from being expressed?
3. If a system is described as a negative repressible operon, what does the word "negative" tell you about the protein involved, and what does "repressible" tell you about its default state?
4. Predict the status of the *lac* repressor protein, the CAP protein, and the overall level of transcription (None, Basal/Low, or High) for the following environmental conditions:
 1. Glucose is highly abundant, and lactose is completely absent from the environment.
 2. Both glucose and lactose are highly abundant in the environment.
 3. Glucose is completely absent, but lactose is highly abundant in the environment.

2.3.3: Eukaryotic Gene Regulation

So far we have explored the flow of genetic information from DNA → RNA → Protein through replication, transcription, RNA processing, and translation. However, an important question remains, if every cell in the body contains essentially the same DNA, why do different cell types look and function so differently? The answer lies in **gene regulation**. Cells control when, where, and how much each gene is expressed. This regulation allows a neuron, muscle cell, and liver cell to perform completely different functions despite sharing the same genome. These differences arise because cells express different subsets of genes. This process is called **differential gene expression**. Gene expression refers to the production of a functional RNA or protein from a gene. Gene regulation refers to the mechanisms that determine when, where, and to what extent a gene is expressed. Some genes are expressed continuously because every cell requires their products. These are known as **constitutive genes** or **housekeeping genes**. Other genes are expressed only under certain conditions or in specific cell types. These are called **regulated genes**.

This chapter looks at the molecular mechanisms that regulate gene expression in eukaryotic cells, from chromatin organization to transcription factors and post-transcriptional regulation. Understanding these mechanisms provides the foundation for developmental biology, cancer genetics, stem cell biology, and many modern genetic technologies.

Epigenetic Gene Regulation

Eukaryotic gene expression is more complex than prokaryotic gene expression because the processes of transcription and translation are physically separated. Unlike prokaryotic cells, eukaryotic cells can regulate gene expression at many different levels, beginning with controlling the access to the DNA. In the baseline state, DNA is wound around positively charged histone proteins to form the structural units called nucleosomes. Because the phosphate backbone of DNA carries a strong negative charge, it binds tightly to unmodified histones. Chromatin can alternate between two configurations: euchromatin, which is loosely packed, open, and transcriptionally active; and heterochromatin, which is highly compacted, closed, and transcriptionally silenced. Alterations to chromatin that are heritable through cell division but do not change the underlying DNA base sequence are defined as **epigenetic modifications**. Transcriptional access to the DNA can be controlled with chromatin remodeling and DNA methylation. Chromatin remodeling changes the way that DNA is associated with histones. DNA methylation is associated with developmental changes and gene silencing.

Epigenetic Control: DNA Methylation

One of the most important mechanisms of epigenetic regulation is **DNA methylation**, the addition of a small chemical group called a **methyl group (-CH₃)** to specific DNA bases in non-coding regions called **CpG islands**. CpG islands are stretches of DNA with a high frequency of cytosine and guanine pairs. In eukaryotic cells, methylation most commonly occurs on the cytosine bases. The addition of methyl groups is catalyzed by enzymes called **DNA methyltransferases (DNMTs)**. These enzymes transfer a methyl group from a donor molecule to the fifth carbon of the cytosine base, producing 5-methylcytosine. Although this chemical modification does not change the nucleotide sequence of DNA, it affects gene expression.

DNA methylation generally silences genes. When methyl groups accumulate within the promoter region of a gene, transcription factors and RNA Polymerase II have a harder time binding to the DNA. As a result, transcription is greatly reduced or completely prevented. In addition to physically interfering with transcription factor binding, methylated DNA attracts specialized proteins that recruit other regulatory complexes that condense chromatin, making the DNA even less accessible to the transcription machinery. The combined effect is that highly methylated genes are often transcriptionally inactive.

DNA methylation does not alter the DNA sequence itself. Instead, it changes how the cell interprets that sequence. For this reason, DNA methylation is considered an epigenetic modification. These methylation tags can last through multiple rounds of mitosis, allowing cells to lock in their specific tissue identities during development. In some instances, these patterns can even bypass embryonic resetting to cross generational lines, leading to **genomic imprinting**, where a gene is transmitted to offspring in a pre-silenced condition depending on which parent it was inherited from.

Epigenetic Control: Histone Modification

The transition between euchromatin and heterochromatin is regulated by adding or removing chemical functional groups directly

onto the histone amino acid tails at the N-terminus of histone proteins. One major modification is **histone acetylation**, a process catalyzed by *histone acetyltransferases (HATs)* that adds acetyl groups to specific positively charged lysine residues. Because acetyl groups carry a neutral charge, this modification masks the positive charge of the histone tail, weakening its attraction to the negatively charged DNA backbone. As a result, the tight grip relaxes, causing the nucleosomes to slide apart and open up the promoter regions so that RNA polymerase and transcription factors can dock to initiate transcription. Once the gene no longer needs to be transcribed, *histone deacetylases (HDACs)* remove these acetyl groups, restoring the positive charge of the histones, causing the DNA to snap tightly back into closed heterochromatin that effectively silences gene expression. Histone acetylation increases gene expression, while deacetylation is associated with a decrease in gene expression.

Histone proteins can also undergo **methylation**, in which one or more methyl groups are added to specific amino acids within histone tails. Unlike DNA methylation, the effects of histone methylation are more complex. Some histone methylation marks promote transcription, while others repress gene expression. The outcome depends on which histone protein is modified, which amino acid is methylated, and the number of methyl groups that are added. For example, methylation of one lysine residue may activate transcription, whereas methylation of another lysine residue may silence the gene.

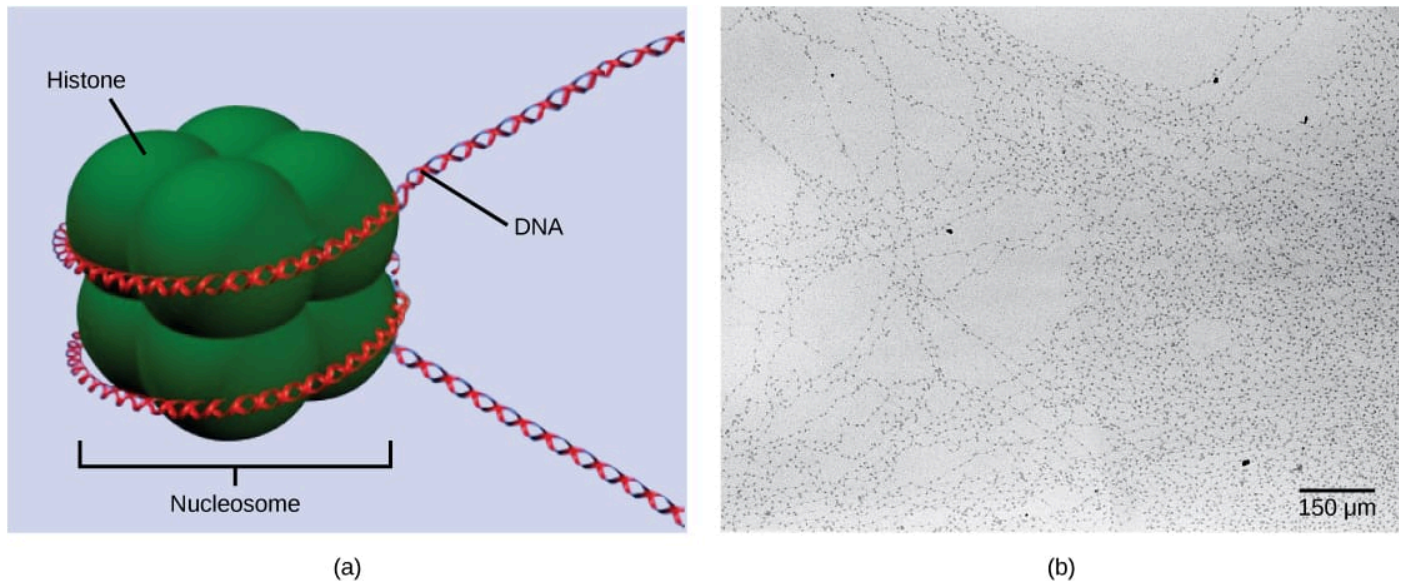


Figure 5.7 DNA is folded around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit “micrograph”: modification of work by Chris Woodcock)

Visual Connection

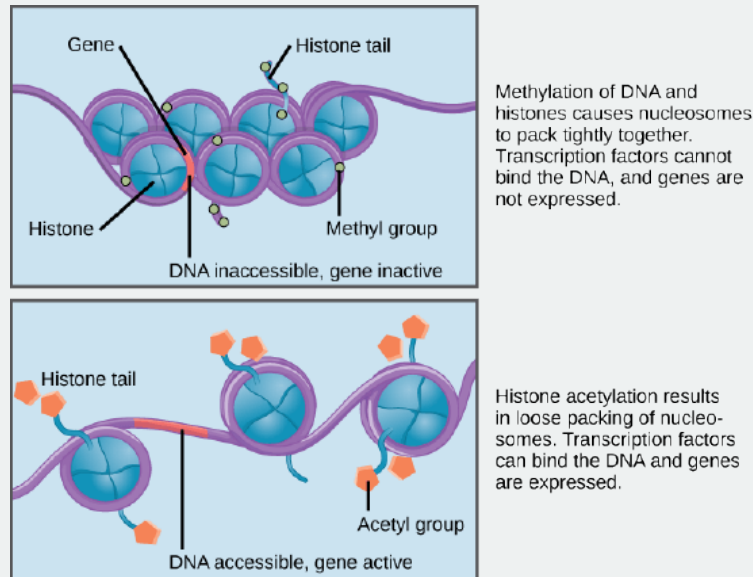


Figure 5.8 Nucleosomes can slide along DNA. When nucleosomes are spaced closely together (top), transcription factors cannot bind and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed. Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

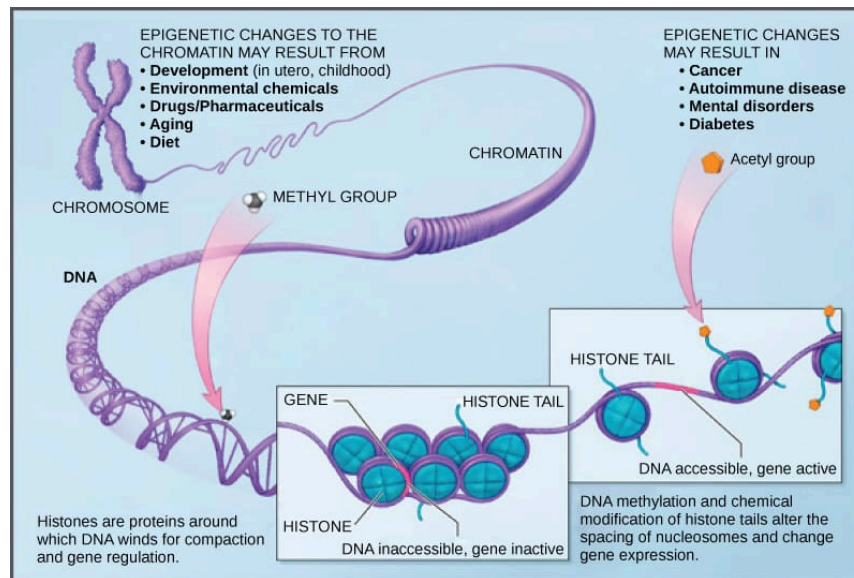


Figure 5.9 Histone proteins and DNA nucleotides can be modified chemically. Modifications affect nucleosome spacing and gene expression. (credit: modification of work by NIH)

Note

[Link to Learning](#)[YouTube Video](#)

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Transcriptional Gene Regulation

Like prokaryotic cells, the transcription of genes in eukaryotes requires the action of an RNA polymerase to bind to a DNA sequence upstream of a gene in order to initiate transcription. However, unlike prokaryotic cells, the eukaryotic RNA polymerase requires other proteins, or *transcription factors*, to facilitate transcription initiation. RNA polymerase by itself cannot initiate transcription in eukaryotic cells. There are two types of transcription factors that regulate eukaryotic transcription: *General (or basal) transcription factors* bind to the core promoter region to assist with the binding of RNA polymerase. *Specific transcription factors* bind to various regions outside of the core promoter region and interact with the proteins at the core promoter to enhance or repress the activity of the polymerase.

The Promoter and the Transcription Machinery

Genes are organized to make the control of gene expression easier. The promoter region is immediately upstream of the coding sequence. This region can be short (only a few nucleotides in length) or long (hundreds of nucleotides long). The longer the promoter, the more available space for proteins to bind. This also adds more control to the transcription process. The length of the promoter is gene-specific and can differ dramatically between genes. Consequently, the level of control of gene expression can also differ quite dramatically between genes. The purpose of the promoter is to bind transcription factors that control the initiation of transcription.

Within the core promoter region, 25 to 35 bases upstream of the transcriptional start site, resides the TATA box. The TATA box has the consensus sequence of 5'-TATAAA-3'. The TATA box is the binding site for a protein complex called TFIID, which contains a TATA-binding protein. Binding of TFIID recruits other transcription factors, including TFIIB, TFIIE, TFIIF, and TFIIH. Some of these transcription factors help to bind the RNA polymerase to the promoter, and others help to activate the transcription initiation complex.

In addition to the TATA box, other binding sites are found in some promoters. Some biologists prefer to restrict the range of the eukaryotic promoter to the core promoter, or polymerase binding site, and refer to these additional sites as **promoter-proximal elements**, because they are usually found within a few hundred base pairs upstream of the transcriptional start site. Examples of these elements are the CAAT box, with the consensus sequence 5'-CCAAT-3' and the GC box, with the consensus sequence 5'-GGGCGG-3'. Specific transcription factors can bind to these promoter-proximal elements to regulate gene transcription. A given gene may have its own combination of these specific transcription-factor binding sites. There are hundreds of transcription factors in a cell, each of which binds specifically to a particular DNA sequence motif. When transcription factors bind to the promoter just upstream of the encoded gene, it is referred to as a **cis-acting element**, because it is on the same chromosome just next to the gene. Transcription factors respond to environmental stimuli that cause the proteins to find their binding sites and initiate transcription of the gene that is needed.

Enhancers and Transcription

In some eukaryotic genes, there are additional regions that help increase or enhance transcription. These regions, called **enhancers**, are not necessarily close to the genes they enhance. They can be located upstream of a gene, within the coding region of the gene, downstream of a gene, or may be thousands of nucleotides away.

Enhancer regions are binding sequences, or sites, for specific transcription factors. When a protein transcription factor binds to its enhancer sequence, the shape of the protein changes, allowing it to interact with proteins at the promoter site. However, since

the enhancer region may be distant from the promoter, the DNA must bend to allow the proteins at the two sites to come into contact. DNA bending proteins help to bend the DNA and bring the enhancer and promoter regions together (Figure 5.10). This shape change allows for the interaction of the specific activator proteins bound to the enhancers with the general transcription factors bound to the promoter region and the RNA polymerase.

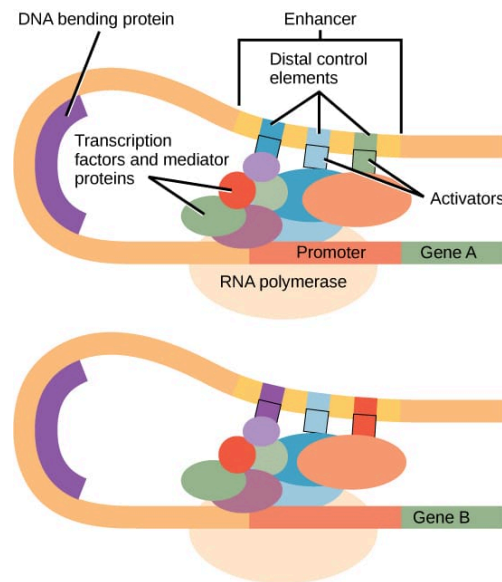


Figure 5.10 Interaction between proteins at the promoter and enhancer sites. An enhancer is a DNA sequence that promotes transcription. Each enhancer is made up of short DNA sequences called distal control elements. Activators bound to the distal control elements interact with mediator proteins and transcription factors. Two different genes may have the same promoter but different distal control elements, enabling differential gene expression.

Turning Genes Off: Transcriptional Repressors

Like prokaryotic cells, eukaryotic cells also have mechanisms to prevent transcription. Transcriptional **repressors** can bind to promoter or enhancer regions and block transcription. Like the transcriptional activators, repressors respond to external stimuli to prevent the binding of activating transcription factors

Post-Transcriptional Gene Regulation

RNA is transcribed, but must be processed into a mature form before translation can begin. This processing that takes place after an RNA molecule has been transcribed, but before it is translated into a protein, is called *post-transcriptional modification*. As with the epigenetic and transcriptional stages of processing, this post-transcriptional step can also be regulated to control gene expression in the cell. If the RNA is not processed, shuttled, or translated, then no protein will be synthesized.

RNA Splicing, the First Stage of Post-transcriptional Control

In eukaryotic cells, the RNA transcript often contains regions, called introns, that are removed prior to translation. The regions of RNA that code for protein are called exons. (Figure 5.11). After an RNA molecule has been transcribed, but prior to its departure from the nucleus to be translated, the RNA is processed and the introns are removed by splicing. Splicing is done by spliceosomes, ribonucleoprotein complexes that can recognize the two ends of the intron, cut the transcript at those two points, and bring the exons together for ligation. Remember from Unit 3, alternative splicing can also occur, where exons are also selectively spliced out to give a different protein isoform.

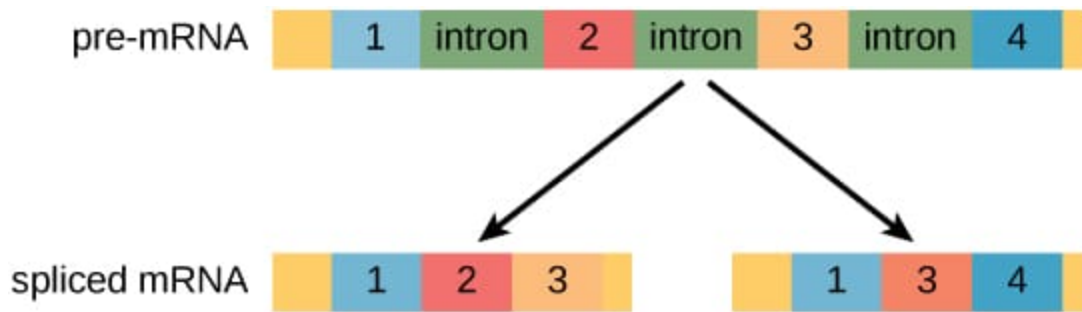


Figure 5.11 Pre-mRNA can be alternatively spliced to create different proteins.

How could alternative splicing evolve? Introns have a beginning- and ending-recognition sequence; it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and instead 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 intron 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 abnormal splicing rather than mutations in a coding sequence. However, alternative splicing could possibly 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.

Question

In the corn snake *Pantherophis guttatus*, there are several different color variants, including amelanistic snakes whose skin patterns display only red and yellow pigments. The cause of amelanism in these snakes was recently identified as the insertion of a transposable element into an intron in the *OCA2* (oculocutaneous albinism) gene. How might the insertion of extra genetic material into an intron lead to a nonfunctional protein?



(left) normal corn snake, (middle) amelanistic yellow corn snake, (right) amelanistic red corn snake

Control of RNA Stability

Before the mRNA leaves the nucleus, it is given two protective "caps" that prevent the ends of the strand from degrading during its journey. 5' and 3' exonucleases can degrade unprotected RNAs. The 5' cap, which is placed on the 5' end of the mRNA, is composed of a methylated guanosine triphosphate molecule (GTP). The GTP is placed "backward" on the 5' end of the mRNA, so that the 5' carbons of the GTP and the terminal nucleotide are linked through three phosphates. The poly-A tail, which is attached to the 3' end, is usually composed of a long chain of adenine nucleotides. These changes protect the two ends of the RNA from exonuclease attack.

Once the RNA is transported to the cytoplasm, the length of time that the RNA resides there can be controlled. Each RNA molecule has a defined lifespan and decays at a specific rate. This rate of decay can influence how much protein is in the cell. If the decay rate is increased, the RNA will not exist in the cytoplasm as long, shortening the time available for translation of the mRNA to occur. Conversely, if the rate of decay is decreased, the mRNA molecule will reside in the cytoplasm longer and more protein can be translated. This rate of decay is referred to as the RNA stability. If the RNA is stable, it will be detected for longer periods of time in the cytoplasm.

Binding of proteins to the RNA can also influence its stability. Proteins called RNA-binding proteins, or RBPs, can bind to the regions of the mRNA just upstream or downstream of the protein-coding region. These regions in the RNA that are not translated into protein are called the **untranslated regions**, or **UTRs**. They are not introns (those have been removed in the nucleus). Rather, these are regions that regulate mRNA localization, stability, and protein translation. The region just before the protein-coding region is called the **5' UTR**, whereas the region after the coding region is called the **3' UTR** (Figure 5.12). The binding of RBPs to these regions can increase or decrease the stability of an RNA molecule, depending on the specific RBP that binds.

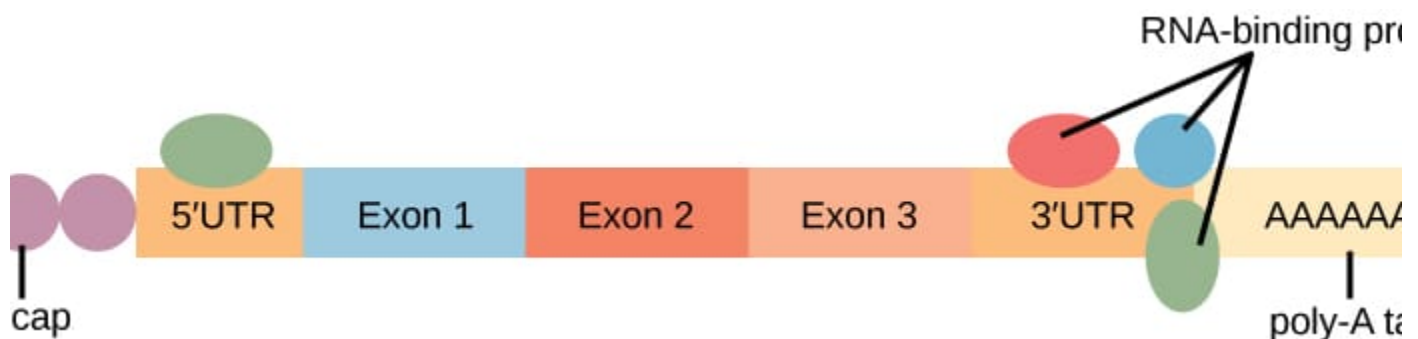


Figure 5.12 RNA-binding proteins. The protein-coding region of this processed mRNA is flanked by 5' and 3' untranslated regions (UTRs). The presence of RNA-binding proteins at the 5' or 3' UTR influences the stability of the RNA molecule.

RNA Stability and microRNAs

In addition to RNA-binding proteins, cells also use a regulatory mechanism known as **RNA interference (RNAi)** to reduce gene expression. RNA interference is a naturally occurring process in which small RNA molecules recognize specific messenger RNAs and prevent them from being translated or promote their degradation. RNA interference regulates gene expression by controlling the fate of mRNA molecules after transcription.

The most common regulators involved in RNAi are **microRNAs (miRNAs)**. MicroRNAs are short, single-stranded RNA molecules approximately 21-24 nucleotides in length. They do not encode proteins. Instead, they function as regulatory RNAs that help determine which messenger RNAs are translated into proteins. MiRNAs are initially synthesized in the nucleus as longer precursor molecules called primary miRNAs that are then processed into shorter pre-miRNAs before being transported into the cytoplasm. Once in the cytoplasm, an enzyme called *Dicer* cleaves the pre-miRNA into a mature double-stranded miRNA. One strand of this molecule becomes incorporated into a protein complex known as the **RNA-induced silencing complex (RISC)**. The RNA component of the RISC base-pairs with complementary sequences on an mRNA and either impede translation of the message or lead to the degradation of the mRNA.

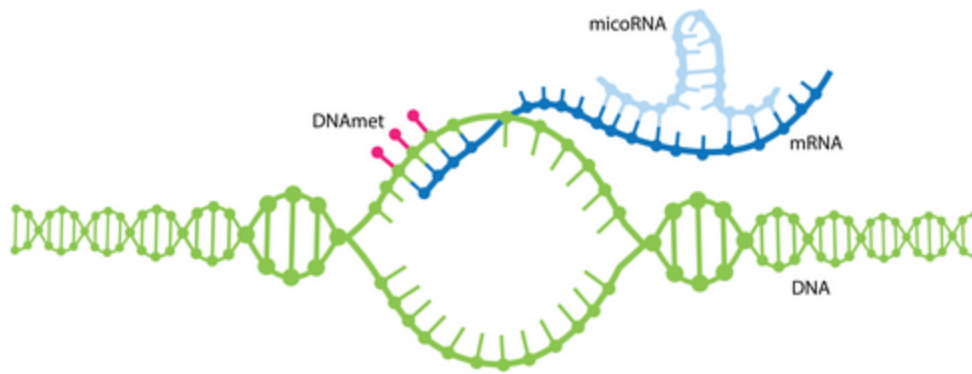


Figure 5.13 An example of how miRNA associates with mRNA. The miRNA is associated with a protein complex that would also be sitting on the mRNA, stopping translation.

Small Interfering RNAs (siRNAs)

Another class of regulatory RNAs involved in RNA interference is the **small interfering RNA (siRNA)**. Like microRNAs, siRNAs are approximately 21–24 nucleotides long and function through the RISC complex. However, siRNAs differ from miRNAs. MicroRNAs are naturally encoded by the genome and usually regulate normal cellular gene expression by binding imperfectly to their target mRNAs. In contrast, siRNAs often arise from double-stranded RNA molecules and usually bind perfectly to their target mRNAs, leading to rapid degradation of the transcript. Because of this high specificity, scientists frequently use synthetic siRNAs as research tools to selectively silence individual genes.

Translational and Post-Translational Control

After RNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. As previously discussed, the stability of the RNA will have a large impact on its translation into a protein. As the stability changes, the amount of time that it is available for translation also changes.

The Initiation Complex and Translation Rate

Like transcription, translation is controlled by proteins that bind and initiate the process. In translation, the complex that assembles to start the process is referred to as the translation initiation complex. In eukaryotes, translation is initiated by binding the initiating met-tRNA_i to the 40S ribosome. This tRNA is brought to the 40S ribosome by a protein initiation factor, eukaryotic initiation factor-2 (eIF-2). The eIF-2 protein binds to the high-energy molecule guanosine triphosphate (GTP). The tRNA-eIF2-GTP complex then binds to the 40S ribosome. A second complex forms on the mRNA. Several different initiation factors recognize the 5' cap of the mRNA and proteins bound to the poly-A tail of the same mRNA, forming the mRNA into a loop. The cap-binding protein eIF4F brings the mRNA complex together with the 40S ribosome complex. The ribosome then scans along the mRNA until it finds a start codon AUG. When the anticodon of the initiator tRNA and the start codon are aligned, the GTP is hydrolyzed, the initiation factors are released, and the large 60S ribosomal subunit binds to form the translation complex. The binding of eIF-2 to the RNA is controlled by **phosphorylation**. If eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP. Therefore, the initiation complex cannot form properly and translation is impeded (Figure 16.14). When eIF-2 remains unphosphorylated, the initiation complex can form normally and translation can proceed.

Visual Connection

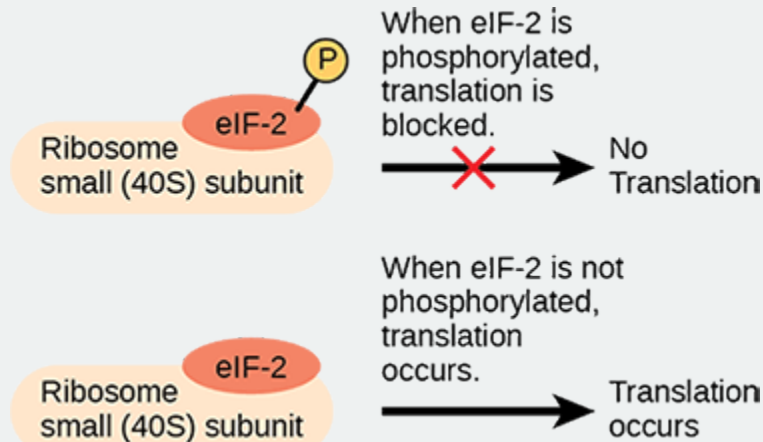


Figure 5.14 Gene expression can be controlled by factors that bind the translation initiation complex.

An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's. What impact do you think this might have on protein synthesis?

Chemical Modifications, Protein Activity, and Longevity

Proteins can be chemically modified with the addition of groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins regulates their activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, in the cytoplasm, or attached to the plasma membrane.

Chemical modifications occur in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. These changes can alter epigenetic accessibility, transcription, mRNA stability, or translation, all resulting in changes in expression of various genes. This is an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment. Because proteins are involved in every stage of gene regulation, the phosphorylation of a protein (depending on the protein that is modified) can alter accessibility to the chromosome, can alter translation (by altering transcription factor binding or function), can change nuclear shuttling (by influencing modifications to the nuclear pore complex), can alter RNA stability (by binding or not binding to the RNA to regulate its stability), can modify translation (increase or decrease), or can change post-translational modifications (add or remove phosphates or other chemical modifications).

The addition of a **ubiquitin** group to a protein marks that protein for degradation. Ubiquitin acts like a flag indicating that the protein lifespan is complete. These proteins are moved to the proteasome, an organelle that functions to remove proteins, to be degraded (Figure 5.15). One way to control gene expression, therefore, is to alter the longevity of the protein.

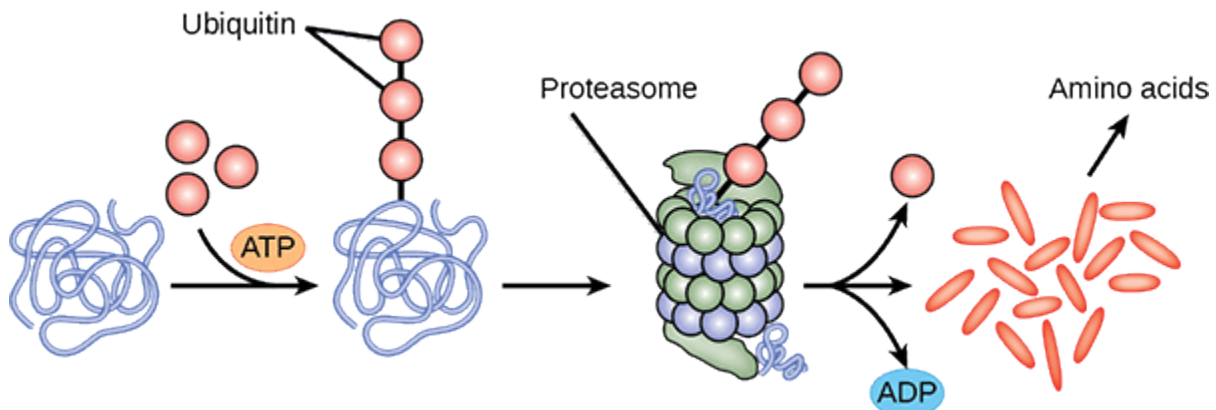


Figure 5.15 Proteins with ubiquitin tags are marked for degradation within the proteasome.

Key Terms

Gene Expression

The process by which the information stored in a gene is used to produce a functional RNA or protein.

Gene Regulation

The control of when, where, and how much a gene is expressed.

Differential Gene Expression

The process by which different cell types express different sets of genes despite having the same DNA.

Constitutive Gene

A gene that is expressed continuously because its product is needed by the cell at all times.

Regulated Gene

A gene whose expression changes in response to developmental signals or environmental conditions.

Chromatin Remodeling

The process of altering chromatin structure to increase or decrease access to DNA.

Epigenetics

Heritable changes in gene expression that occur without altering the DNA sequence.

DNA Methylation

The addition of methyl groups to DNA, often reducing gene expression.

Histone Acetylation

The addition of acetyl groups to histone proteins, generally making chromatin more accessible for transcription.

Histone Methylation

A chemical modification of histones that can either activate or repress gene expression depending on the location of the modification.

Activator

A transcription factor that increases gene expression.

Repressor

A transcription factor that decreases gene expression.

Enhancer

A regulatory DNA sequence that increases transcription by binding activator proteins.

Silencer

A regulatory DNA sequence that decreases transcription by binding repressor proteins.

RNA Interference (RNAi)

A mechanism that uses small RNA molecules to reduce or silence gene expression.

microRNA (miRNA)

A small RNA molecule that regulates gene expression by reducing translation or promoting mRNA degradation.

Small Interfering RNA (siRNA)

A short RNA molecule that directs the degradation of complementary mRNA molecules.

Check Your Understanding

1. If nearly every cell in your body contains the same DNA, why do nerve cells and muscle cells have different structures and functions?

2. What is the difference between gene expression and gene regulation?
3. Distinguish between constitutive and regulated genes.
4. Why is gene regulation essential for multicellular organisms?
5. Compare euchromatin and heterochromatin.
6. Why must chromatin be remodeled before many genes can be transcribed?
7. What is epigenetics?
8. How does DNA methylation generally affect gene expression?
9. How does histone acetylation influence chromatin structure?
10. Explain why epigenetic changes do not alter the DNA sequence.
11. Compare activators and repressors.
12. Why are transcription factors considered essential regulators of gene expression?
13. What is an enhancer?
14. What is a silencer?
15. How can an enhancer regulate a gene that is thousands of base pairs away?
16. Why is mRNA stability important for gene expression?
17. How can cells regulate protein production after transcription has already occurred?
18. Why is post-transcriptional regulation especially important in eukaryotic cells?
19. What is RNA interference?
20. If a mutation completely deactivated a cell's *histone deacetylase (HDAC)* enzymes, what would you expect to happen to the structural state of that cell's chromatin (Euchromatin vs. Heterochromatin)?

2.3.4: Cancer and Gene Regulation

Cancer is not a single disease but includes many different diseases. In cancer cells, mutations modify cell-cycle control and cells don't stop growing as they normally would. Mutations can also alter the growth rate or the progression of the cell through the cell cycle. One example of a gene modification that alters the growth rate is increased phosphorylation of cyclin B, a protein that controls the progression of a cell through the cell cycle and serves as a cell-cycle checkpoint protein.

For cells to move through each phase of the cell cycle, the cell must pass through checkpoints. This ensures that the cell has properly completed the step and has not encountered any mutation that will alter its function. Many proteins, including cyclin B, control these checkpoints. The phosphorylation of cyclin B, a post-translational event, alters its function. As a result, cells can progress through the cell cycle unimpeded, even if mutations exist in the cell and its growth should be terminated. This post-translational change of cyclin B prevents it from controlling the cell cycle and contributes to the development of cancer.

Cancer: Disease of Altered Gene Expression

Cancer can be described as a disease of altered gene expression. There are many proteins that are turned on or off (gene activation or gene silencing) that dramatically alter the overall activity of the cell. A gene that is not normally expressed in that cell can be switched on and expressed at high levels. This can be the result of gene mutation or changes in gene regulation (epigenetic, transcription, post-transcription, translation, or post-translation).

Changes in epigenetic regulation, transcription, RNA stability, protein translation, and post-translational control can be detected in cancer. While these changes don't occur simultaneously in one cancer, changes at each of these levels can be detected when observing cancer at different sites in different individuals. Therefore, changes in histone acetylation (epigenetic modification that leads to gene expression), activation of transcription factors by phosphorylation, increased RNA stability, increased translational control, and protein modification can all be detected at some point in various cancer cells. Scientists are working to understand the common changes that give rise to certain types of cancer or how a modification might be exploited to destroy a tumor cell.

Tumor Suppressor Genes, Oncogenes, and Cancer

In normal cells, some genes function to prevent excess, inappropriate cell growth. These are tumor-suppressor genes, which are active in normal cells to prevent uncontrolled cell growth. There are many tumor-suppressor genes in cells. The most studied tumor-suppressor gene is p53, which is mutated in over 50 percent of all cancer types. The p53 protein itself functions as a transcription factor. It can bind to sites in the promoters of genes to initiate transcription. Therefore, the mutation of p53 in cancer will dramatically alter the transcriptional activity of its target genes.

Link to Learning

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



Proto-oncogenes are positive cell-cycle regulators. When mutated, proto-oncogenes can become oncogenes and cause cancer. Overexpression of the oncogene can lead to uncontrolled cell growth. This is because oncogenes can alter transcriptional activity, stability, or protein translation of another gene that directly or indirectly controls cell growth. An example of an oncogene involved in cancer is a protein called myc. Myc is a transcription factor that is aberrantly activated in Burkett's Lymphoma, a cancer of the lymph system. Overexpression of myc transforms normal B cells into cancerous cells that continue to grow uncontrollably. High B-cell numbers can result in tumors that can interfere with normal bodily function. Patients with Burkett's lymphoma can develop

tumors on their jaw or in their mouth that interfere with the ability to eat.

Cancer and Epigenetic Alterations

Silencing genes through epigenetic mechanisms is also very common in cancer cells. There are characteristic modifications to histone proteins and DNA that are associated with silenced genes. In cancer cells, the DNA in the promoter region of silenced genes is methylated on cytosine DNA residues in CpG islands. Histone proteins that surround that region lack the acetylation modification that is present when the genes are expressed in normal cells. This combination of DNA methylation and histone deacetylation (epigenetic modifications that lead to gene silencing) is commonly found in cancer. When these modifications occur, the gene present in that chromosomal region is silenced. Increasingly, scientists understand how epigenetic changes are altered in cancer. Because these changes are temporary and can be reversed—for example, by preventing the action of the histone deacetylase protein that removes acetyl groups, or by DNA methyl transferase enzymes that add methyl groups to cytosines in DNA—it is possible to design new drugs and new therapies to take advantage of the reversible nature of these processes. Indeed, many researchers are testing how a silenced gene can be switched back on in a cancer cell to help re-establish normal growth patterns.

Genes involved in the development of many other illnesses, ranging from allergies to inflammation to autism, are thought to be regulated by epigenetic mechanisms. As our knowledge of how genes are controlled deepens, new ways to treat diseases like cancer will emerge.

Cancer and Transcriptional Control

Alterations in cells that give rise to cancer can affect the transcriptional control of gene expression. Mutations that activate transcription factors, such as increased phosphorylation, can increase the binding of a transcription factor to its binding site in a promoter. This could lead to increased transcriptional activation of that gene that results in modified cell growth. Alternatively, a mutation in the DNA of a promoter or enhancer region can increase the binding ability of a transcription factor. This could also lead to the increased transcription and aberrant gene expression that is seen in cancer cells.

Researchers have been investigating how to control the transcriptional activation of gene expression in cancer. Identifying how a transcription factor binds, or a pathway that activates where a gene can be turned off, has led to new drugs and new ways to treat cancer. In breast cancer, for example, many proteins are overexpressed. This can lead to increased phosphorylation of key transcription factors that increase transcription. One such example is the overexpression of the epidermal growth-factor receptor (EGFR) in a subset of breast cancers. The EGFR pathway activates many protein kinases that, in turn, activate many transcription factors which control genes involved in cell growth. New drugs that prevent the activation of EGFR have been developed and are used to treat these cancers.

Cancer and Post-transcriptional Control

Changes in the post-transcriptional control of a gene can also result in cancer. Recently, several groups of researchers have shown that specific cancers have altered expression of miRNAs. Because miRNAs bind to the 3' UTR of RNA molecules to degrade them, overexpression of these miRNAs could be detrimental to normal cellular activity. Too many miRNAs could dramatically decrease the RNA population, leading to a decrease in protein expression. Several studies have demonstrated a change in the miRNA population in specific cancer types. It appears that the subset of miRNAs expressed in breast cancer cells is quite different from the subset expressed in lung cancer cells or even from normal breast cells. This suggests that alterations in miRNA activity can contribute to the growth of breast cancer cells. These types of studies also suggest that if some miRNAs are specifically expressed only in cancer cells, they could be potential drug targets. It would, therefore, be conceivable that new drugs that turn off miRNA expression in cancer could be an effective method to treat cancer.

Cancer and Translational/Post-translational Control

There are many examples of how translational or post-translational modifications of proteins arise in cancer. Modifications are found in cancer cells from the increased translation of a protein to changes in protein phosphorylation to alternative splice variants of a protein. An example of how the expression of an alternative form of a protein can have dramatically different outcomes is seen in colon cancer cells. The c-Flip protein, a protein involved in mediating the cell-death pathway, comes in two forms: long (c-FLIPL) and short (c-FLIPS). Both forms appear to be involved in initiating controlled cell-death mechanisms in

normal cells. However, in colon cancer cells, expression of the long form results in increased cell growth instead of cell death. Clearly, the expression of the wrong protein dramatically alters cell function and contributes to the development of cancer.

New Drugs to Combat Cancer: Targeted Therapies

Scientists are using what is known about the regulation of gene expression in disease states, including cancer, to develop new ways to treat and prevent disease development. Many scientists are designing drugs on the basis of the gene expression patterns within individual tumors. This idea, that therapy and medicines can be tailored to an individual, has given rise to the field of personalized medicine. With an increased understanding of gene regulation and gene function, medicines can be designed to specifically target diseased cells without harming healthy cells. Some new medicines, called targeted therapies, have exploited the overexpression of a specific protein or the mutation of a gene to develop a new medication to treat disease. One such example is the use of anti-EGF receptor medications to treat the subset of breast cancer tumors that have very high levels of the EGF protein. Undoubtedly, more targeted therapies will be developed as scientists learn more about how gene expression changes can cause cancer.

2.4: MUTATIONS AND DNA REPAIR

2.4.1: DNA Damage and Repair

2.4.2: Types of Mutations

2.4.1: DNA Damage and Repair

Every time a cell divides, it must faithfully copy its entire genome. At the same time, DNA molecules are continuously exposed to factors that can damage or alter their structure. Despite this, most cells maintain an astonishingly low mutation rate, ensuring that genetic information is accurately passed from one generation of cells to the next. The ability to preserve the integrity of the genome is known as **genome stability**. Genome stability is essential for normal growth, development, and survival. If DNA damage accumulates unchecked, cells may lose important genetic information, produce defective proteins, or begin dividing uncontrollably, potentially leading to diseases such as cancer.

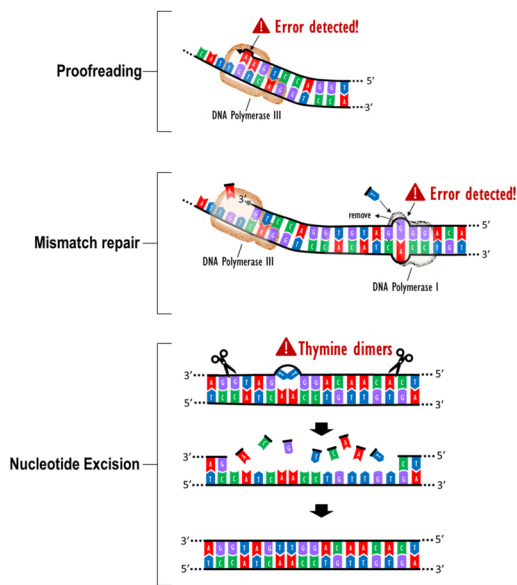
Sources of DNA Damage

The DNA sequence of an organism is an active chemical structure continuously challenged by both internal processes and external environmental factors. Each day, an individual cell experiences thousands of spontaneous DNA lesions arising from two primary categories of cellular threats. Endogenous damage originates directly from normal, unavoidable internal processes. These include spontaneous replication errors, interactions with reactive oxygen species produced during cellular respiration, the spontaneous loss of nitrogenous bases known as depurination, and the accidental chemical modification of bases via deamination. Exogenous damage is caused by environmental factors, called **mutagens**, operating outside the cell. Common environmental mutagens include ultraviolet (UV) radiation from sunlight, ionizing radiation from X-rays and gamma rays, tobacco smoke, industrial toxins, viruses, and pollutants. If these combined endogenous and exogenous lesions were left unaddressed, mutations would compound exponentially.

Polymerase Proofreading

The initial precision of DNA polymerase relies heavily on strict complementary base pairing and is immediately reinforced by the enzyme's structural capacity for self-correction. The first active line of defense against permanent mutation occurs in real-time during synthesis through proofreading. Occasionally, an incorrect nucleotide slips into place, such as an adenine pairing incorrectly with a cytosine. Because mismatched bases cannot hydrogen-bond correctly, they create a distinct physical distortion in the shape of the DNA double helix that prevents the mismatched pair from fitting properly within the active site of the enzyme.

When the polymerase detects this localized structural distortion, it pauses forward replication and activates its secondary 3' → 5' exonuclease site, which cleanly clips out the incorrect nucleotide. Once the error is excised, normal DNA synthesis resumes. Proofreading improves overall replication accuracy roughly 100-fold. For the rare replication errors that manage to slip past the polymerase unnoticed, cells use a secondary post-replication quality control system known as **mismatch repair (MMR)**. Operating immediately after DNA replication is complete, specialized mismatch repair proteins scan the newly synthesized double helix to locate mispaired bases. Once a mismatch is identified, the repair machinery differentiates the newly synthesized strand from the parental template, cuts out the segment containing the error, and recruits DNA polymerase to fill the gap with correct nucleotides before DNA ligase seals the remaining nick.



6.1 Different DNA repair mechanisms.

DNA Repair

DNA lesions also occur long after replication has finished, so cells must have different specialized repair pathways tailored to distinct types of structural damage. When cellular metabolism or spontaneous chemical decay alters only a single nucleotide (such as an oxidized base, a deaminated cytosine, or a site experiencing depurination) cells use the **base excision repair (BER)** pathway. During this process, specialized enzymes called DNA glycosylases recognize and cleanly snip out the damaged nitrogenous base, leaving behind an empty sugar-phosphate backbone. Secondary enzymes then remove the remaining sugar fragment, allowing DNA polymerase to insert the correct replacement nucleotide and DNA ligase to seal the strand back together.

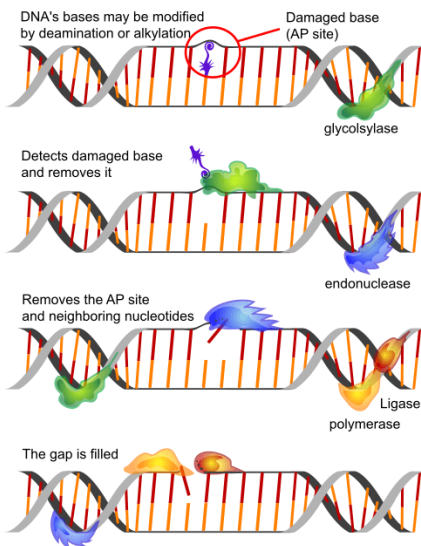


Figure 6.2 The base excision repair pathway.

For larger, bulky lesions that cause distortions to the DNA double helix, cells use the **nucleotide excision repair (NER)** pathway. This mechanism is particularly handy for removing thymine (or pyrimidine) dimers, which are aberrant covalent bonds between adjacent bases induced by ultraviolet radiation (Figure 6.3). Instead of removing a single isolated base, the nucleotide excision repair machinery cuts out a short, multi-nucleotide stretch of the damaged strand surrounding the lesion. DNA polymerase then uses the intact opposing strand as a template to synthesize a fresh replacement sequence, which is then locked into place by

DNA ligase.

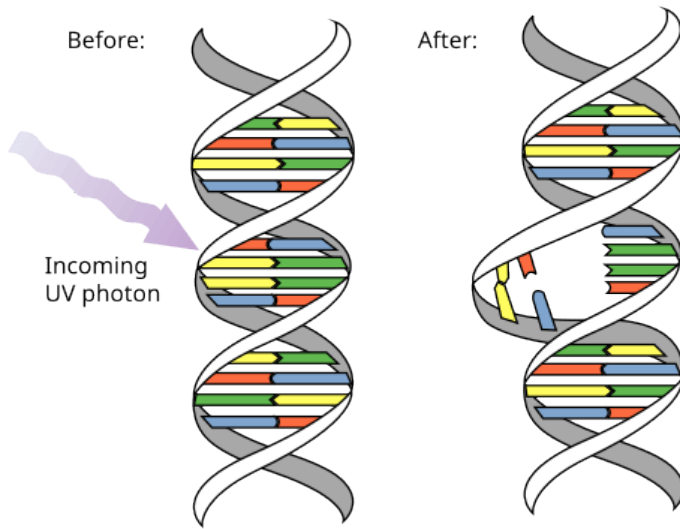


Figure 6.3 Ultraviolet (UV) harm the DNA molecules in different ways. In one common damage event, adjacent bases bond with each other, instead of across the “ladder.” This makes a bulge, and the distorted DNA molecule does not function properly.

The biggest threats to genome stability are double-strand breaks, which occur when both strands of the DNA double helix are severed simultaneously by ionizing radiation, harsh chemicals, replication fork collapse, or mechanical chromosomal stress. Unrepaired double-strand breaks lead to catastrophic chromosomal rearrangements or immediate cell death, forcing cells to choose between two emergency repair strategies. The first mechanism, **homologous recombination**, relies on using an identical, undamaged sister chromatid as a physical template to guide a highly accurate repair of the broken strands. When a matching template is unavailable, cells must rely on **non-homologous end joining (NHEJ)**, a faster process that directly glues the broken DNA ends back together. While non-homologous end joining successfully preserves overall chromosome integrity, it lacks a template and is highly error-prone, frequently introducing small insertions or deletions at the splice site.

Key Terms

Genome Stability

The maintenance of an accurate and functional DNA sequence through faithful DNA replication and efficient DNA repair mechanisms.

DNA Damage

Any alteration to the chemical structure of DNA that can interfere with DNA replication, transcription, or chromosome integrity.

Mutagen

A physical or chemical agent that increases the frequency of mutations by damaging DNA.

Proofreading

The ability of DNA polymerase to detect and remove incorrectly paired nucleotides during DNA replication before DNA synthesis continues.

3' → 5' Exonuclease Activity

The proofreading function of DNA polymerase that removes incorrectly incorporated nucleotides from the growing DNA strand.

Mismatch Repair (MMR)

A DNA repair pathway that corrects mismatched base pairs remaining after DNA replication has been completed.

Base Excision Repair (BER)

A DNA repair pathway that removes and replaces damaged or altered individual nucleotides, such as oxidized or deaminated

bases.

DNA Glycosylase

An enzyme that recognizes and removes damaged bases during base excision repair.

Nucleotide Excision Repair (NER)

A DNA repair pathway that removes short stretches of DNA containing bulky lesions, such as thymine dimers, and replaces them with newly synthesized DNA.

Thymine Dimer

A DNA lesion formed when two adjacent thymine bases become covalently linked following exposure to ultraviolet (UV) radiation, distorting the DNA double helix.

Double-Strand Break

A type of DNA damage in which both strands of the DNA double helix are broken.

Homologous Recombination (HR)

An accurate DNA repair pathway that repairs double-strand breaks using an undamaged homologous DNA molecule as a template.

Non-Homologous End Joining (NHEJ)

A DNA repair pathway that directly joins broken DNA ends without using a template. Although rapid, it is more likely to introduce small mutations than homologous recombination.

Check Your Understanding

1. What is a mutagen, and how can mutagens increase the frequency of mutations?
2. Why is DNA damage considered unavoidable in living cells?
3. What is the function of the 3' → 5' exonuclease activity of DNA polymerase?
4. How does mismatch repair differ from DNA polymerase proofreading?
5. Why is mismatch repair considered a second line of defense against mutations?
6. Why do cells require multiple DNA repair pathways instead of one universal repair system?
7. What types of DNA damage are repaired by base excision repair?
8. What role do DNA glycosylases play during base excision repair?
9. What are thymine dimers, and how are they repaired?
10. Why are double-strand breaks considered one of the most dangerous forms of DNA damage?
11. Compare homologous recombination and non-homologous end joining.

This page titled [2.4.1: DNA Damage and Repair](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Jennifer Yeack.

2.4.2: Types of Mutations

Despite the remarkable accuracy of DNA replication and the many repair systems that maintain genome stability, some changes in DNA escape detection. When these changes become permanently incorporated into the genome, they are known as **mutations**. A mutation is a permanent change in the nucleotide sequence of the DNA. Mutations may occur spontaneously during normal cellular processes or be induced by environmental agents such as ultraviolet radiation, chemicals, or ionizing radiation. Mutations are a normal part of biology. Every organism accumulates mutations throughout its lifetime, and every generation inherits new mutations from its parents. While some mutations are harmful and contribute to disease, many have little or no effect, and a small number provide beneficial traits that can be favored by natural selection.

Base Substitutions: Transitions versus Transversions

The simplest type of mutation is the **point mutation**, which occurs when a single nucleotide or base pair is altered within the DNA sequence. The most common point mutations involve base substitutions where one specific nucleotide is swapped out for another. At the chemical level, these substitutions are further divided into transition and transversion mutations based on the nucleotide type. A **transition mutation** occurs when a purine is replaced by another purine (such as an adenine swapping with a guanine) or when a pyrimidine is replaced by another pyrimidine (such as a cytosine swapping with a thymine). **Transversion mutations** occur when a double-ringed purine is replaced by a single-ringed pyrimidine, or vice versa, such as an adenine switching to a thymine or a guanine changing to a cytosine. Because transitions replace bases with chemically and structurally similar counterparts, they occur with much higher frequency in nature. Transversions, by contrast, force a mismatch between bases of fundamentally different sizes, which produces a much larger physical distortion in the geometry of the DNA double helix.

Although base substitutions occur strictly at the DNA level, their true biological significance is determined by how they alter the downstream messenger RNA transcript and the ultimate protein encoded by the gene. A **silent mutation** changes the codon to another triplet that still specifies the exact same amino acid; for example, changing a GAA codon to GAG results in no physical change to the protein because both triplets code for glutamic acid. These mutations are entirely harmless and are made possible because the genetic code is degenerate.

A **missense mutation** occurs when a base substitution transforms a codon so that it specifies a completely different amino acid, such as changing GAG into GTG, which swaps a glutamic acid for a valine. The consequences of a missense mutation vary; some have negligible effects if the replacement amino acid shares similar chemical properties with the original, while others drastically alter the protein's tertiary structure and ruin its function. This destructive potential is illustrated by sickle cell disease, a severe inherited blood disorder caused by a single missense mutation in the β -globin gene.

The most severe point mutation outcome is a **nonsense mutation**, which converts a standard amino acid codon into a premature stop codon. During translation, the ribosome interprets this accidental stop signal as the end of the gene, stopping protein synthesis prematurely. The resulting truncated proteins are usually unstable, short-lived, and completely nonfunctional, which is why nonsense mutations generally devastating.

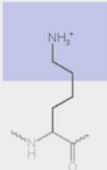
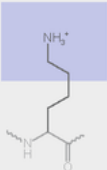
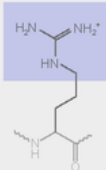
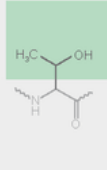
Point mutations					
No mutation	Silent		Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr
					
				basic	polar

Figure 6.7 Examples of point mutations.

Insertions, Deletions, and Reading Frames

Genetic mutations can also involve the physical addition or removal of nucleotides rather than simple substitutions. An **insertion** occurs when one or more extra nucleotides are added into the DNA sequence, whereas a **deletion** involves the permanent removal of one or more base pairs. Because protein-coding genes are read continuously in strict groups of three nucleotides, the impact of an insertion or deletion depends entirely on whether the number of affected nucleotides is divisible by three.

If an insertion or deletion involves a number of nucleotides that is not a multiple of three, it causes a **frameshift mutation**. This shifts the ribosome's reading frame downstream of the mutation site, scrambling every single subsequent codon. Consequently, frameshift mutations produce entirely wrong amino acid sequences and routinely introduce premature stop codons, making them some of the most destructive mutations a gene can experience.

A variation of these length changes is the **trinucleotide repeat expansion**, where a short, three-nucleotide sequence naturally present in a gene is abnormally duplicated over and over again. As the number of these repeats expands past a critical threshold, it disrupts normal gene expression and protein folding. This specific mechanism is responsible for several inherited neurological conditions, most notably Huntington disease. Unlike standard point mutations, repeat expansions are highly dynamic; they can continuously increase in size as they are passed from one generation to the next, causing clinical symptoms to appear earlier and manifest with greater severity in successive generations of an affected family.

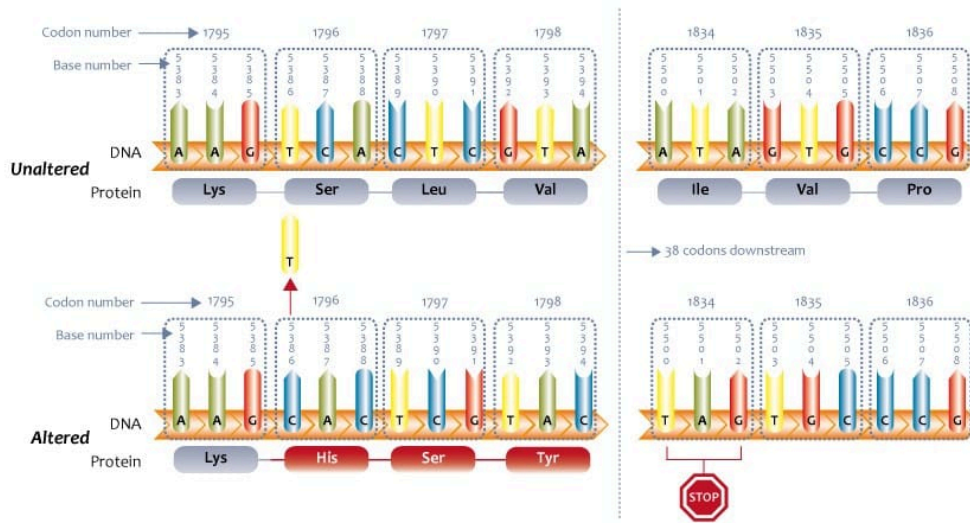


Figure 6.8 Example of a frameshift mutation.

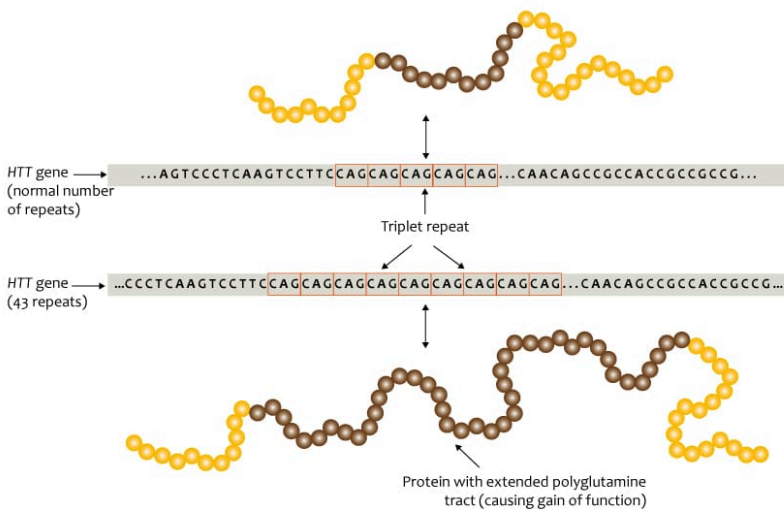


Figure 6.9 Example of trinucleotide repeat.

Key Terms

Mutation

A permanent change in the nucleotide sequence of DNA.

Spontaneous Mutation

A mutation that occurs naturally through normal cellular processes.

Induced Mutation

A mutation caused by exposure to environmental mutagens.

Point Mutation

A mutation affecting a single nucleotide or base pair.

Substitution Mutation

A mutation in which one nucleotide is replaced by another.

Transition Mutation

A substitution of one purine for another purine or one pyrimidine for another pyrimidine.

Transversion Mutation

A substitution of a purine for a pyrimidine or vice versa.

Silent Mutation

A mutation that does not alter the amino acid sequence of a protein.

Missense Mutation

A mutation that changes one amino acid in a protein.

Nonsense Mutation

A mutation that converts an amino acid codon into a stop codon.

Insertion

The addition of one or more nucleotides to a DNA sequence.

Deletion

The removal of one or more nucleotides from a DNA sequence.

Frameshift Mutation

A mutation caused by an insertion or deletion that alters the reading frame of a gene.

Trinucleotide Repeat Expansion

A mutation in which a three-nucleotide sequence is repeated more times than normal.

Check Your Understanding

1. What is a mutation, and how does it differ from DNA damage?
2. Compare spontaneous and induced mutations. Give two examples of each.
3. What is a point mutation?
4. What is the difference between a transition and a transversion mutation?
5. Explain why silent mutations often have little or no effect on protein function.
6. Compare missense and nonsense mutations. Which is generally more likely to produce a nonfunctional protein, and why?
7. What is a frameshift mutation, and why are frameshift mutations often more severe than point mutations?
8. How do insertions and deletions differ from substitution mutations?
9. What are trinucleotide repeat expansions, and what genetic disorder is commonly associated with them?
10. Why do some mutations have no detectable effect on an organism?

“Mutations and Health” by [U.S. National Library of Medicine](#) is in the [Public Domain](#)

3: MODULE 3

3.1: Inheritance

3.1.1: Mendel's Experiments

3.1.2: Laws of Inheritance

3.1.3: Extensions of the Laws of Inheritance

3.1.4: Black fur color - a dominant trait

3.1.5: Yellow fur color - a recessive trait

3.1.6: Brindle color - partial dominance and epistasis

3.1.7: White spotting - when there's more than two alleles

3.1.8: KIT - embryonic lethality

3.1.9: Hemophilia - a sex-linked disorder

3.1.10: It's not all in the genes - the effect of environment

3.1.11: Pleiotropy - one gene affects more than one trait

3.1: INHERITANCE

3.1.1: Mendel's Experiments

3.1.2: Laws of Inheritance

3.1.3: Extensions of the Laws of Inheritance

3.1.4: Black fur color - a dominant trait

3.1.5: Yellow fur color - a recessive trait

3.1.6: Brindle color - partial dominance and epistasis

3.1.7: White spotting - when there's more than two alleles

3.1.8: KIT - embryonic lethality

3.1.9: Hemophilia - a sex-linked disorder

3.1.10: It's not all in the genes - the effect of environment

3.1.11: Pleiotropy - one gene affects more than one trait

3.1.1: Mendel's Experiments

Johann Gregor Mendel (1822–1884) (Figure 8.1) 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.



Figure 8.1: Johann Gregor Mendel set the framework for the study of genetics.

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 **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 8.2). 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 **F₁**, or the first filial (filial = daughter or son), generation. Once Mendel examined the characteristics in the F₁ generation of plants, he allowed them to self-fertilize naturally. He then collected and grew the seeds from the F₁ plants to produce the **F₂**, or second filial, generation. Mendel's experiments extended beyond the F₂ generation to the F₃ generation, F₄ generation, and so on, but it was the ratio of characteristics in the P, F₁, and F₂ generations that were the most intriguing and became the basis of Mendel's postulates.

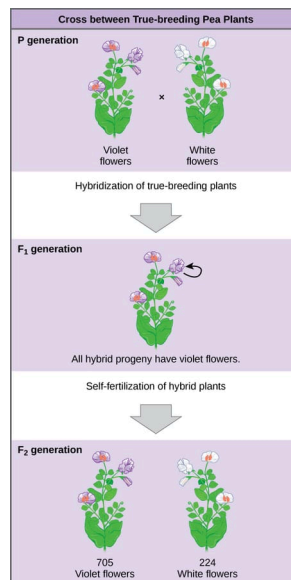


Figure 8.2 Mendel's process for performing crosses included examining flower color.

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 F₁ and F₂ plants and reported results from thousands of F₂ 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 F₁ 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 F₁ generation.

Importantly, Mendel did not stop his experimentation there. He allowed the F₁ plants to self-fertilize and found that 705 plants in the F₂ 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 8.3).

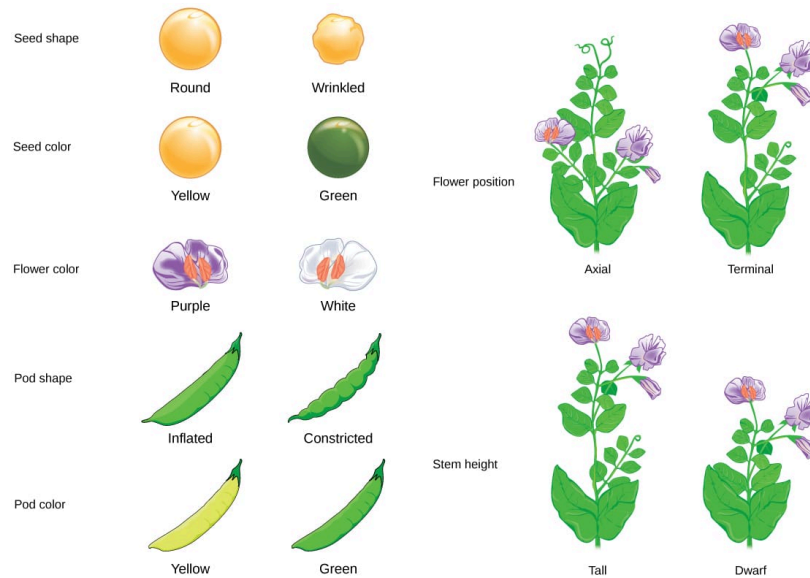


Figure 8.3 Mendel identified seven pea plant characteristics.

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.

Key Terms

Heredity

The transmission of genetic information and traits from parents to offspring.

True-Breeding (Pure-Breeding)

Describes organisms that consistently produce offspring with the same trait when they self-fertilize or breed with individuals of the same phenotype.

Hybridization (Cross)

The mating of two individuals with different traits to study inheritance patterns.

Parental Generation (P Generation)

The original pair of individuals used in a genetic cross.

First Filial Generation (F_1 Generation)

The offspring produced from a cross between the parental generation.

Second Filial Generation (F_2 Generation)

The offspring produced by crossing or self-fertilizing individuals from the F_1 generation.

Trait

A heritable characteristic or feature of an organism.

Continuous Variation

Variation in a trait that exhibits a continuous range of values, such as human height.

Discontinuous Variation

Variation in which individuals exhibit distinct, easily categorized traits with few intermediate forms.

Dominant Trait

The trait that is expressed in an organism when at least one dominant allele is present.

Recessive Trait

A trait that is expressed only when an organism possesses two recessive alleles.

Reciprocal Cross

A pair of genetic crosses in which the male and female phenotypes are reversed to determine whether inheritance depends on the sex of the parent.

Check Your Understanding

1. Why were Mendel's discoveries not immediately accepted by the scientific community?
2. What is the difference between continuous variation and discontinuous variation?
3. Why was discontinuous variation particularly useful for Mendel's experiments?
4. Why was the garden pea (*Pisum sativum*) an excellent model organism for studying inheritance?
5. What does it mean for a plant to be true-breeding?
6. Describe how Mendel performed a hybridization between two pea plants.
7. What are the P, F_1 , and F_2 generations?
8. Why did Mendel analyze thousands of offspring rather than only a few plants?
9. What happened to the recessive trait in the F_1 generation?
10. Why was the reappearance of the recessive trait in the F_2 generation important?
11. What phenotypic ratio did Mendel commonly observe in the F_2 generation?
12. What is a reciprocal cross, and why did Mendel perform one?
13. How did Mendel's observations contradict the blending theory of inheritance?
14. What is the difference between a dominant trait and a recessive trait?

Footnotes

1. 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 www.mendelweb.org/Mendel.plain.html]

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [3.1.1: Mendel's Experiments](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [8.1: Mendel's Experiments](#) by OpenStax is licensed [CC BY 4.0](#).

3.1.2: Laws of Inheritance

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, F₁, and F₂ 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 F₁ 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 F₂ offspring (Figure 8.4). Therefore, the F₁ 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 F₁ yellow plants that received a Y allele from their yellow parent and a y allele from their green parent had the genotype Yy.

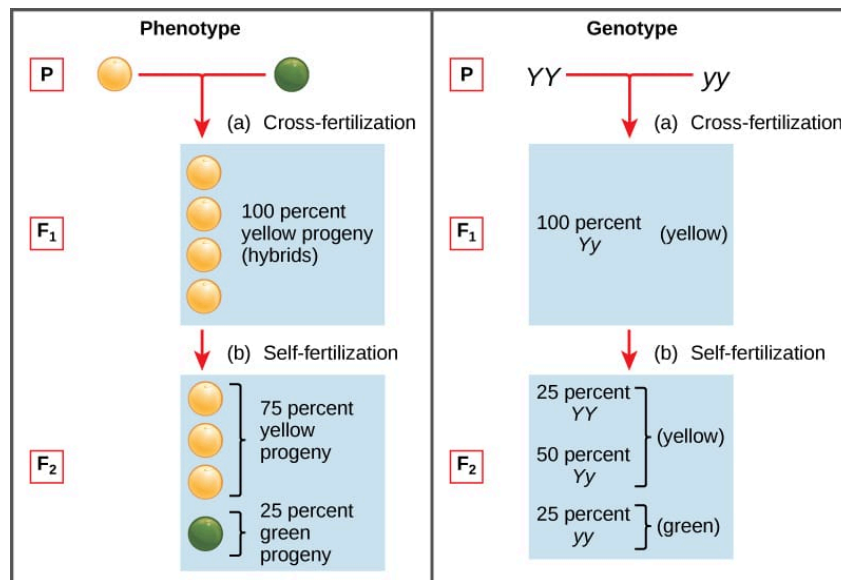


Figure 8.4 Phenotypes are physical expressions of traits that are transmitted by alleles. Capital letters represent dominant alleles and lowercase letters represent recessive alleles. The phenotypic ratios are the ratios of visible characteristics. The genotypic ratios are the ratios of gene combinations 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 F₁ heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant, and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosomes. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype), and the recessive allele will only be observed in homozygous recessive individuals (Table 8.1).

Table 8.1 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 8.5), and these offspring will breed true when self-crossed.



Figure 8.5 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 8.7) 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 8.6).

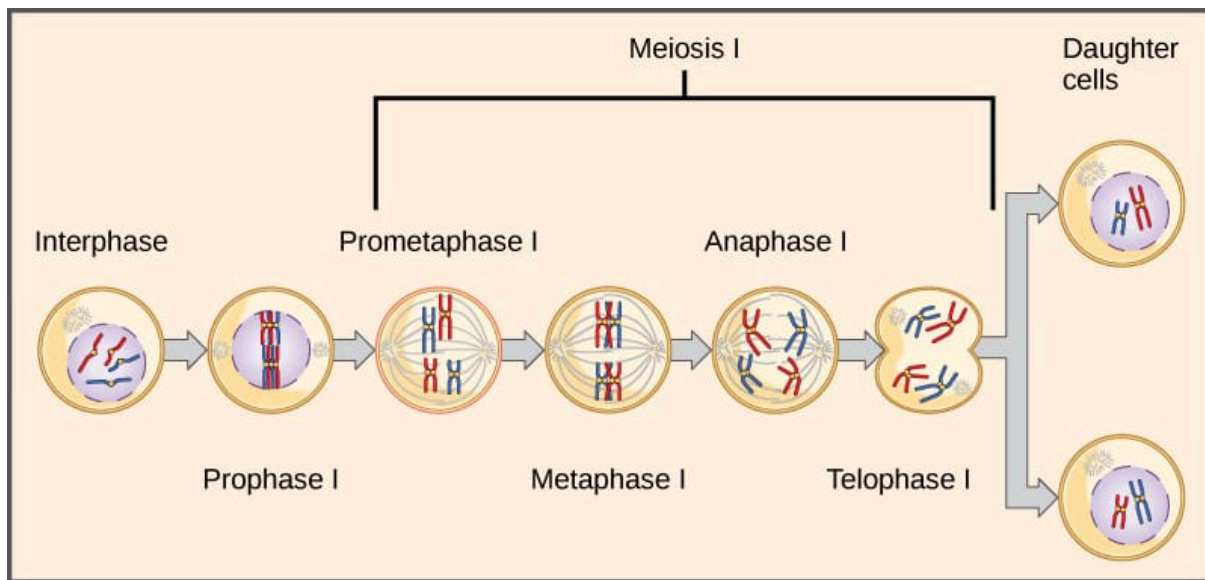


Figure 8.6 The first division in meiosis is shown.

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 8.7). Alternatively, if the dominant-expressing organism is a heterozygote, the F_1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes. The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

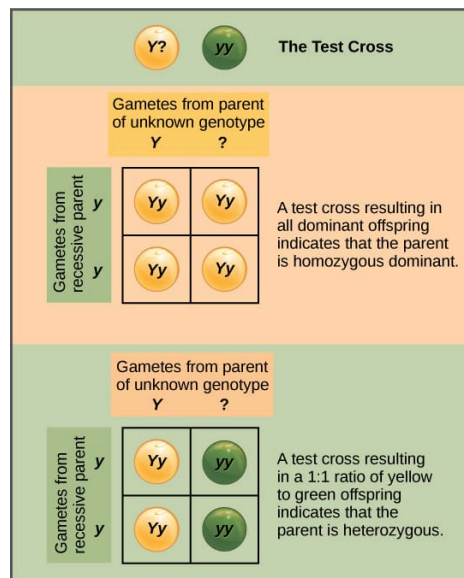


Figure 8.7 A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

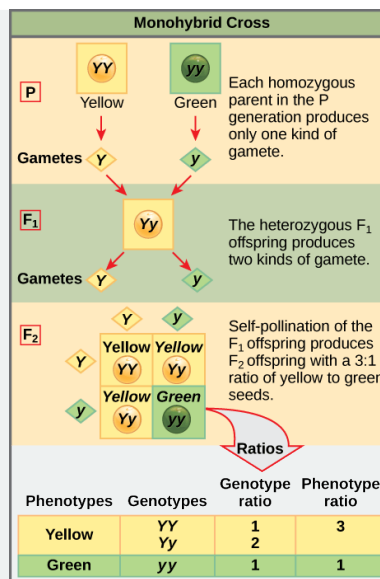


Figure 8.8 This Punnett square shows the cross between plants with yellow seeds and green seeds. The cross between the true-breeding P plants produces F_1 heterozygotes that can be self-fertilized. The self-cross of the F_1 generation can be analyzed with a Punnett square to predict the genotypes of the F_2 generation. Given an inheritance pattern of dominant-recessive, the genotypic and phenotypic ratios can then be determined.

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?

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 8.9).

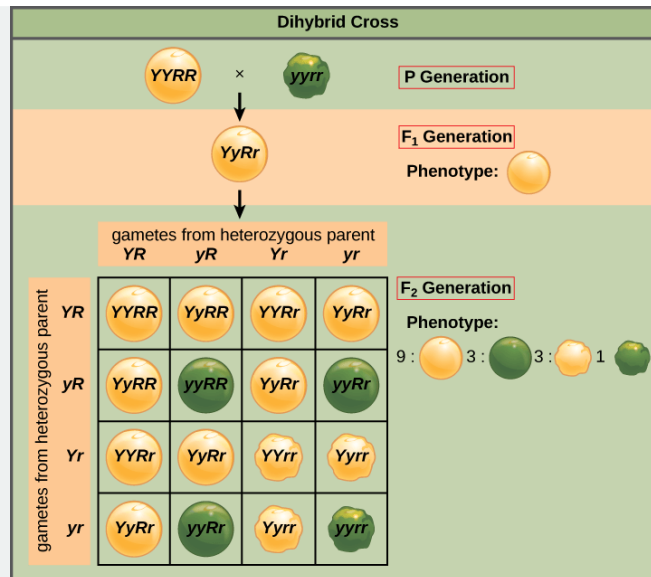


Figure 8.9 A dihybrid cross in pea plants involves the genes for seed color and texture. The P cross produces F₁ offspring that are all heterozygous for both characteristics. The resulting 9:3:3:1 F₂ phenotypic ratio is obtained using a Punnett square.

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 gametes produced by the F₁ individuals must have one allele from each of the two genes. For example, a gamete could get an R allele for the seed shape gene and either a Y or a y allele for the seed color gene. It cannot get both an R and an r allele; each gamete can have only one allele per gene. The law of independent assortment states that a gamete into which an r allele is sorted would be equally likely to contain either a Y or a y allele. Thus, there are four equally likely gametes that can be formed when the $RrYy$ heterozygote is self-crossed, as follows: RY , rY , Ry , and ry . Arranging these gametes along the top and left of a 4 × 4 Punnett square (Figure 8.9) gives us 16 equally likely genotypic combinations. From these genotypes, we find a phenotypic ratio of **9 round-yellow:3 round-green:3 wrinkled-yellow:1 wrinkled-green** (Figure 8.9). 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 8.10).

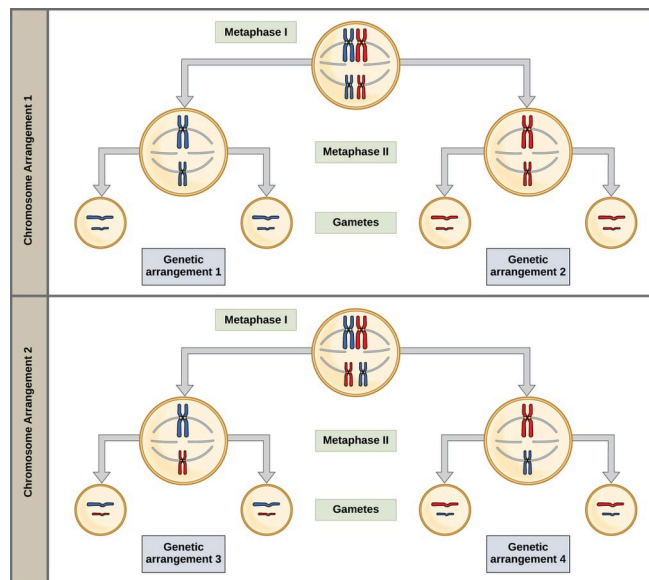


Figure 8.10 The random segregation into daughter nuclei that happens during the first division in meiosis can lead to a variety of possible genetic arrangements.

Key Terms

Gene

A unit of heredity consisting of a DNA sequence that influences a specific trait.

Allele

An alternative version of a gene found at the same location (locus) on homologous chromosomes.

Phenotype

The observable physical or biochemical characteristics of an organism.

Genotype

The genetic makeup of an organism, consisting of the alleles it possesses.

Homozygous

Having two identical alleles for a particular gene.

Heterozygous

Having two different alleles for a particular gene.

Dominant Allele

An allele that is expressed in the phenotype when at least one copy is present.

Recessive Allele

An allele that is expressed only when two copies are present.

Law of Dominance

Mendel's principle stating that in a heterozygous individual, the dominant allele masks the expression of the recessive allele.

Monohybrid Cross

A genetic cross involving one gene or one characteristic.

Punnett Square

A diagram used to predict the possible genotypes and phenotypes of offspring from a genetic cross.

Probability

The mathematical likelihood that a particular event will occur.

Genotypic Ratio

The ratio of the different genotypes expected among offspring from a genetic cross.

Phenotypic Ratio

The ratio of the different observable traits expected among offspring from a genetic cross.

Law of Segregation

Mendel's first law, stating that the two alleles for a gene separate during gamete formation so that each gamete receives only one allele.

Test Cross

A cross between an individual displaying the dominant phenotype and a homozygous recessive individual to determine the unknown genotype.

Dihybrid Cross

A genetic cross involving two different genes or characteristics.

Law of Independent Assortment

Mendel's second law, stating that alleles of different genes assort independently into gametes during meiosis, provided the genes are not genetically linked.

Independent Assortment

The random distribution of homologous chromosome pairs during meiosis I, producing different combinations of maternal and paternal chromosomes in gametes.

Check Your Understanding

1. What is the difference between a gene and an allele?
2. Explain the difference between genotype and phenotype.
3. Compare homozygous and heterozygous genotypes.
4. Why can two organisms have the same phenotype but different genotypes?
5. What is the difference between a dominant allele and a recessive allele?
6. State Mendel's Law of Dominance.
7. Why does a recessive trait disappear in the F_1 generation but reappear in the F_2 generation?
8. Can a heterozygous individual express a recessive phenotype? Explain your answer.
9. What is a monohybrid cross?
10. What information does a Punnett square provide?
11. A heterozygous tall pea plant (Tt) is crossed with a homozygous recessive short plant (tt). Predict the genotypic and phenotypic ratios of the offspring.
12. What is the difference between a genotypic ratio and a phenotypic ratio?
13. Why are Punnett squares based on probabilities rather than certainties?
14. State Mendel's Law of Segregation.
15. How does meiosis provide the biological basis for the Law of Segregation?
16. Why does each gamete contain only one allele for each gene?
17. Explain why Mendel consistently observed a 3:1 phenotypic ratio in the F_2 generation of monohybrid crosses.
18. What is the purpose of a test cross?
19. Why is a homozygous recessive individual always used in a test cross?
20. How can the results of a test cross reveal whether an individual with a dominant phenotype is homozygous or heterozygous?
21. What is a dihybrid cross?
22. State Mendel's Law of Independent Assortment.
23. How does meiosis explain the Law of Independent Assortment?
24. Why can a dihybrid heterozygote produce four different types of gametes?
25. Explain how the classic 9:3:3:1 phenotypic ratio is produced in a dihybrid cross.
26. Under what circumstances does the Law of Independent Assortment not apply?

Comprehensive Chapter Review

1. Explain the relationship between genes, alleles, genotype, and phenotype.
2. Compare homozygous and heterozygous genotypes.
3. Describe how Mendel's Law of Dominance explains the appearance of the F_1 generation.
4. Explain how the Law of Segregation accounts for the 3:1 phenotypic ratio observed in monohybrid crosses.
5. Construct a Punnett square for a cross between two heterozygous individuals ($Aa \times Aa$) and determine the expected genotypic and phenotypic ratios.
6. Explain why Punnett squares predict probabilities rather than the exact outcomes of individual crosses.
7. Describe the purpose of a test cross and explain how it can distinguish between homozygous dominant and heterozygous individuals.
8. Compare monohybrid and dihybrid crosses.
9. Explain how independent assortment increases genetic variation among offspring.
10. Why is the orientation of homologous chromosome pairs during metaphase I essential for Mendel's Law of Independent Assortment?
11. A plant with genotype **TtRr** is crossed with another **TtRr** plant.
 - How many different gametes can each parent produce?
 - What phenotypic ratio would you expect among the offspring?
 - Which Mendelian law explains this ratio?
12. Explain how the processes of meiosis provide the physical basis for both the **Law of Segregation** and the **Law of Independent Assortment**.

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [3.1.2: Laws of Inheritance](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [8.2: Laws of Inheritance](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

3.1.3: Extensions of the Laws of 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 8.11), a cross between a homozygous parent with white flowers ($C^W C^W$) and a homozygous parent with red flowers ($C^R C^R$) will produce offspring with pink flowers ($C^R C^W$). (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 $C^R C^R$:2 $C^R C^W$:1 $C^W C^W$, 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 8.11 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 ($I^A I^A$ and $I^B I^B$) express either the A or the B phenotype, and heterozygotes ($I^A I^B$) express both phenotypes equally. The $I^A I^B$ 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 8.12).

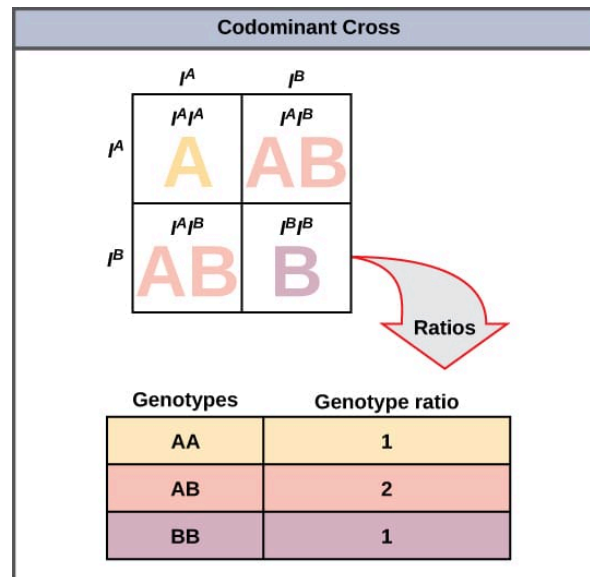


Figure 8.12 This Punnet square shows an AB/AB blood type cross.

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level, such that many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype in the natural population as the **wild type** (often abbreviated "+"). All other phenotypes or genotypes are considered **variants** (mutants) of this typical form, meaning they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is the ABO blood-type system in humans. In this case, there are three alleles circulating in the population. The I^A allele codes for A molecules on the red blood cells, the I^B allele codes for B molecules on the surface of red blood cells, and the i allele codes for no molecules on the red blood cells. In this case, the I^A and I^B alleles are codominant with each other and are both dominant over the i allele. Although there are three alleles present in a population, each individual only gets two of the alleles from their parents. This produces the genotypes and phenotypes shown in Figure 8.13. 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.

Inheritance of the ABO Blood System in Humans			
	I^A	I^B	i
I^A	$I^A I^A$ A	$I^A I^B$ AB	$I^A i$ A
I^B	$I^B I^A$ AB	$I^B I^B$ B	$I^B i$ B
i	$i I^A$ A	$i I^B$ B	$i i$ O

Figure 8.13 Inheritance of the ABO blood system in humans is shown.

EVOLUTION IN ACTION: Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae*, and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly. When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1 percent. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most effective malarial treatments can vary by geographic region.

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it is infective to humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait.

In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that comes about because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions in which this drug is widely used as an over-the-counter malaria remedy. As is common with pathogens that multiply to large numbers within an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide malaria burden.¹

Sex-Linked Traits

In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes—one pair of *non-homologous* chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or **autosomes**. In addition to 22 homologous pairs of autosomes, human females have a homologous pair of X chromosomes, whereas human males have an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains fewer genes. When a gene being examined is present on the X, but not the Y, chromosome, it is **X-linked**.

Eye color in *Drosophila*, the common fruit fly, was the first X-linked trait to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX. In flies the wild-type eye color is red (X^W) and is dominant to white eye color (X^w) (Figure 8.14). 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. Hemizyosity 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 8.14 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 8.15). 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).

ART CONNECTION

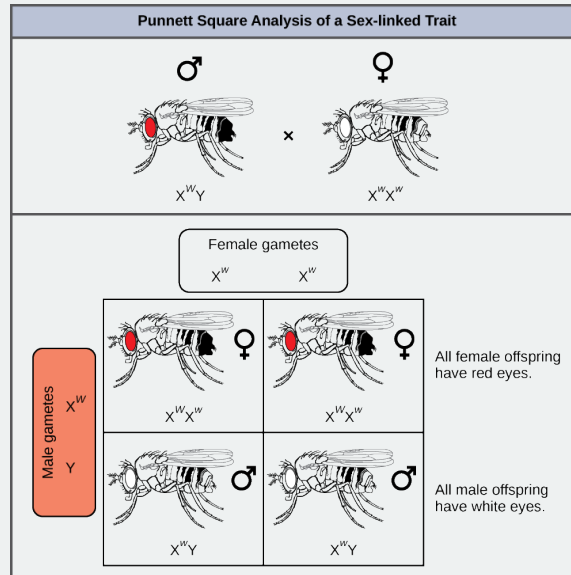


Figure 8.15 Crosses involving sex-linked traits often give rise to different phenotypes for the different sexes of offspring, as is the case for this cross involving red and white eye color in *Drosophila*. In the diagram, w is the white-eye mutant allele and W is the wild-type, red-eye allele.

What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

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

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

[View this YouTube Video online](#)



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. 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 crossing over, 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 8.16). 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.

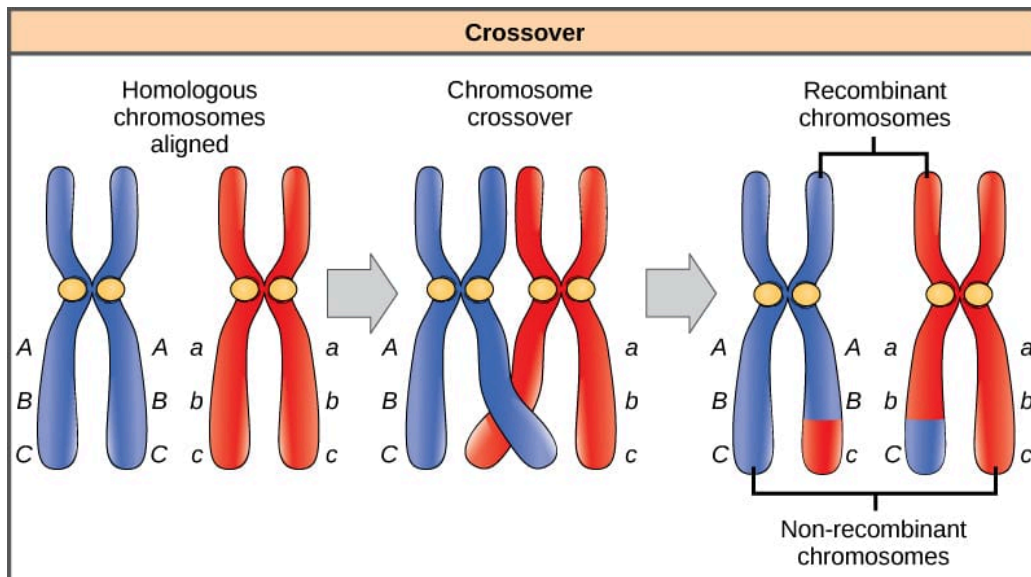


Figure 8.16 The process of crossover, or recombination, occurs when two homologous chromosomes align and exchange a segment of genetic material.

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 8.17). Therefore, the genotypes $AAcc$, $Aacc$, and $aacc$ all produce the same albino phenotype. A cross between heterozygotes for both genes ($AaCc \times AaCc$) would generate offspring with a phenotypic ratio of 9 agouti:3 black:4 albino (Figure 8.17). In this case, the C gene is epistatic to the A gene.

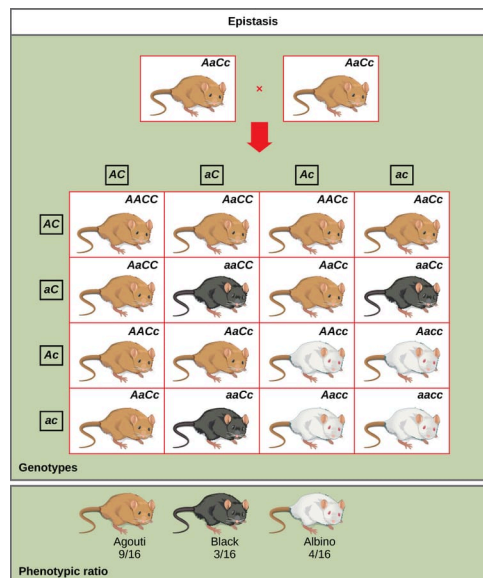


Figure 8.17 In this example of epistasis, one gene (C) masks the expression of another (A) for coat color. When the C allele is present, coat color is expressed; when it is absent (cc), no coat color is expressed. Coat color depends on the A gene, which shows dominance, with the recessive homozygote showing a different phenotype than the heterozygote or dominant homozygote.

Key Terms

Incomplete Dominance

A pattern of inheritance in which the heterozygous phenotype is intermediate between the two homozygous phenotypes.

Codominance

A pattern of inheritance in which both alleles are fully and simultaneously expressed in the heterozygous individual.

Multiple Alleles

The existence of more than two alleles for a gene within a population, although each individual possesses only two alleles.

Wild-Type Allele

The most common allele or phenotype found in a natural population.

Carrier

A heterozygous individual who possesses a recessive allele but does not express the recessive phenotype.

ABO Blood Group System

A human blood type system controlled by three alleles (I^A , I^B , and i), demonstrating both codominance and multiple alleles.

Sex-Linked Trait

A trait determined by a gene located on a sex chromosome, most commonly the X chromosome.

X-Linked Inheritance

A pattern of inheritance involving genes located on the X chromosome.

Hemizygous

Having only one copy of a gene rather than the typical pair. Human males are hemizygous for most X-linked genes.

Carrier Female

A heterozygous female who carries an X-linked recessive allele but usually does not express the trait.

Linked Genes

Genes located close together on the same chromosome that tend to be inherited together.

Linkage Map

A chromosome map showing the relative positions of genes based on recombination frequencies.

Polygenic Inheritance

Inheritance in which multiple genes contribute to a single trait.

Epistasis

An interaction between genes in which one gene masks or modifies the expression of another gene.

Epistatic Gene

The gene whose alleles mask or suppress the expression of another gene.

Hypostatic Gene

The gene whose expression is masked by an epistatic gene.

Check Your Understanding

1. Why are incomplete dominance, codominance, and multiple alleles considered extensions of Mendel's work rather than exceptions to his laws?
2. Compare incomplete dominance and codominance.
3. In a cross between two pink snapdragons ($C^R C^W \times C^R C^W$), what genotypic and phenotypic ratios would you expect?
4. Why does the ABO blood group system demonstrate codominance?
5. What are multiple alleles?
6. Why can a population have more than two alleles for a gene while an individual has only two?
7. List the six possible genotypes of the ABO blood group system.
8. Explain why both I^A and I^B are dominant over i but codominant with each other.
9. What is a sex-linked trait?
10. Why are X-linked recessive disorders more common in males than in females?
11. What does it mean for a male to be hemizygous?
12. Why do reciprocal crosses involving X-linked genes often produce different results?
13. Explain why a carrier female may not express an X-linked recessive disorder.
14. What are linked genes?
15. Why do linked genes violate Mendel's Law of Independent Assortment?
16. What is a recombinant gamete?
17. Why are genes located farther apart on a chromosome more likely to recombine?
18. What is polygenic inheritance?
19. Why do polygenic traits often show continuous variation?
20. What is epistasis?
21. Compare an epistatic gene and a hypostatic gene.
22. Explain how epistasis differs from simple dominant-recessive inheritance.
23. A woman with blood type **AB** has children with a man who has blood type **O**.
 - What blood types are possible among their children?
 - What are the expected proportions of each blood type?
24. A carrier female for an X-linked recessive disorder has children with an unaffected male.
 - What proportion of their sons would be expected to have the disorder?
 - What proportion of their daughters would be expected to be carriers?

Footnotes

1. Sumiti Vinayak et al., "Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*," *PLoS Pathogens* 6 (2010): e1000830.

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [3.1.3: Extensions of the Laws of Inheritance](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [8.3: Extensions of the Laws of Inheritance](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

3.1.4: Black fur color - a dominant trait



Figure 8.18 This chocolate lab has two recessive alleles of the TYRP1 gene. (Credit: [Rob Hanson](#); photo from [Wikimedia](#).)

Most of us are familiar with the labrador retriever dog breed, such as the chocolate lab seen in Figure 8.18. But have you ever thought about what makes this dog brown? The difference between brown and black coat color in dogs is caused by a mutation in the TYRP1 gene. The TYRP1 gene provides instructions for making an enzyme called tyrosinase-related protein 1. This enzyme is required to produce a pigment called eumelanin. Eumelanin is a dark colored pigment. The TYRP1 gene is located on chromosome 11 in dogs (Parker, 2001).

A group of scientists who were interested in determining what caused the difference between black and brown coats sequenced the DNA within the protein-coding region of the TYRP1 gene (Schmutz, 2002). They identified three variations in the DNA making up the TYRP1 gene between brown dogs and black dogs. These variations in DNA sequence are examples of different **alleles** of the TYRP1 gene.

Table 3.1.4.1: Variations in the TYRP1 allele that lead to brown color in dogs. Data from Schmutz, 2002.

Location	Black DNA sequence	Brown DNA sequence	Effect on protein
exon 2	TGT	CGT	changes a cysteine amino acid to a serine
exon 5	CAG	TAG	introduces a premature stop codon which results in 330 amino acids instead of 512 amino acids in the protein
exon 5	CCT	— (deleted)	deletion of a proline amino acid

All of these variations in the DNA sequence are predicted to cause a change in the amino acid sequence of the TYRP1 protein. These changes affect the production of eumelanin pigment, which is black in color. When eumelanin is not being produced correctly, the dog appears brown instead of black.

Like other diploid organisms, dogs all have two copies of the TYRP1 gene (one from their male parent, one from their female parent). Dogs that are homozygous for the black allele (dogs that have two copies of the black allele) are obviously going to be black in color. Dogs that are homozygous for the brown allele are obviously going to be brown. Dogs that are heterozygous (dogs that have one black allele and one brown allele) appear black. The black and brown colors do not blend together: the black allele covers up the brown allele. This means that the black allele is **dominant** over the brown allele. Remember that dominant alleles cover up **recessive** alleles. If there is one dominant allele present, the dog will appear black. The brown allele is recessive to the black allele. There must be two copies of the recessive brown allele present in order for the dog to appear brown.



Figure 8.19 Black and brown phenotypes in labrador retrievers. (Credit: [demealliffe](#); from [Wikimedia](#))

Remember that genotypes can be abbreviated with a single letter and that the letter which is chosen is typically the first letter of the dominant trait. In this case, the letter “B” is used to represent the dominant black allele, while “b” represents a recessive brown allele.

The reason that the black allele is dominant over the brown allele in this specific situation is because the black allele produces functional TYRP1 protein, while the brown allele does not. The presence of one functional allele produces enough TYRP1 protein allows the cells to produce eumelanin and appear black.

Remember: dominant does not mean “better” or “more normal”. Black color does not confer any special advantages on dogs compared to brown color. It’s just a difference.



Figure 8.20 What alleles of TYRP1 does this black lab puppy have? We can’t tell by looking at it. The puppy could be homozygous (BB) or heterozygous (Bb). Since black is completely dominant over brown, both options would be black. (Credit: [Alice Birkin](#))

Let’s visualize the inheritance of black and brown using a **pedigree**. The pedigree in Figure 8.21 shows a litter of puppies. The shaded symbol shows a brown puppy, while open symbols are black individuals.

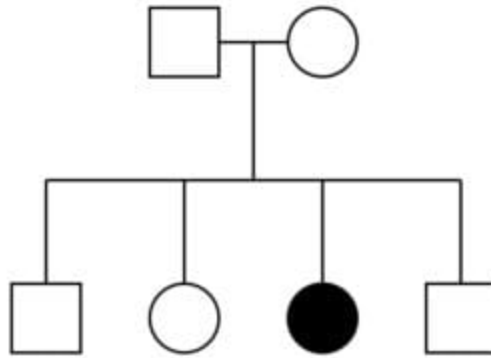


Figure 8.21 An example litter of puppies. The filled-in symbol shows a brown individual.

To interpret this pedigree, let's start with information that we already know:

- Brown is recessive, which means brown individuals must have the genotype bb . In this pedigree, brown individuals are filled in.
- Black is dominant, which means black individuals must have at least one B allele. Their genotype could be either BB or Bb . In this pedigree, black individuals are not filled in.

Figure 8.22 shows the same pedigree, but with information about the individual's genotypes filled in.

1. The shaded individual, who is a brown female puppy, must have the genotype bb . If she had any B alleles, she would be black because the black allele is dominant over the brown allele.
2. In order for the brown puppy to have the genotype bb , she must have gotten two " b " alleles: one from each of her parents. We know that her parents are both black (because they are unshaded), which means they must have at least one " B " allele. This means that both parents must be heterozygous: Bb .
3. The three black puppies must have at least one " B " allele in order for them to be black in color. However, we can't tell whether they are homozygous dominant (BB) or heterozygous (Bb) since both of those genotypes would result in black color. One way to represent this on a pedigree is $B-$, meaning that the second allele could be either B or b .

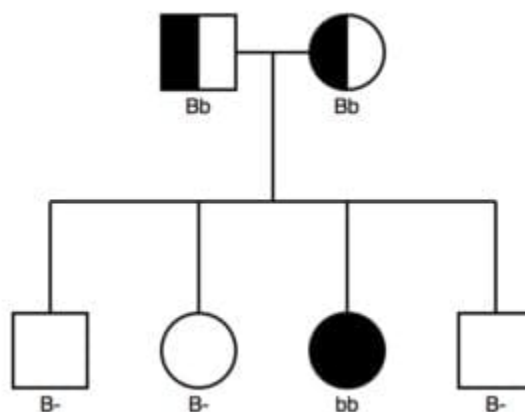


Figure 8.22 Genotypes of the individuals in this pedigree.

We can also show the cross between these parents as a Punnett square (Figure 8.23). We would expect $1/4$ of the offspring to have the genotype bb , and that is what we see in the pedigree above.

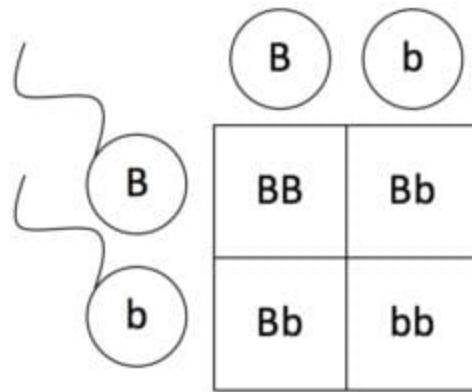


Figure 8.23 The information from the pedigree shown in Figure 3.1.4.5 can also be shown as a Punnett square.

Human Connection

A small number of mutations in the TYRP1 gene have been found to cause oculocutaneous albinism type 3. This condition includes a form of albinism called rufous oculocutaneous albinism, which has been described primarily in dark-skinned people from southern Africa. Affected individuals have reddish-brown skin, ginger or red hair, and hazel or brown irises. Two TYRP1 mutations are known to cause this form of albinism in individuals from Africa. One mutation replaces a protein building block (amino acid) in tyrosine-related protein 1 with a signal that prematurely stops protein production. This mutation, written as Ser166Ter or S166X, affects the amino acid serine at protein position 166. The other mutation, written as 368delA, deletes a single DNA building block from the TYRP1 gene. Other alterations in this gene have been reported in a few affected people of non-African heritage. Most TYRP1 mutations lead to the production of an abnormally short, nonfunctional version of tyrosinase-related protein 1. Because this enzyme plays a role in normal pigmentation, its loss leads to the changes in skin, hair, and eye coloration that are characteristic of oculocutaneous albinism.



Photo credit: [Muntuwandi](#); from [Wikipedia](#).

References

Unless otherwise noted, text and images by Lisa Bartee, 2016.

Parker HG, Yuhua X, Mellersh CS, Khan S, Shibuya H, Johnson GS, Ostrander EA. Sept 2001. [Meiotic linkage mapping of 52 genes onto the canine map does not identify significant levels of microrearrangement](#). Mamm Genome. 12(9):713-8.

Schmutz SM, Berryere TG, Goldfinch AD. 2002. [TYRP1 and MC1R genotypes and their effects on coat color in dogs](#). Mammalian Genome 13, 380-387.

OpenStax, Biology. OpenStax CNX. May 27, 2016 <http://cnx.org/contents/s8Hh0oOc@9.1...ls-Experiments>

OpenStax, Biology. OpenStax CNX. May 27, 2016 <http://cnx.org/contents/s8Hh0oOc@9.1...of-Inheritance>

Information about TYRP1 and oculocutaneous albinism type 3: “Tyrop1” by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

This page titled [3.1.4: Black fur color - a dominant trait](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Lisa Bartee, Walter Shriner and Catherine Creech.

3.1.5: Yellow fur color - a recessive trait

Yellow color in dogs

Labrador retrievers don't only come in brown and black, they also come in yellow. Yellow color in labs is caused by variations in a different gene: MC1R. This gene controls the production of the melanocortin 1 receptor protein. MC1R is located on chromosome 5 in dogs (Schmutz, 2001).



Figure 8.24 This yellow lab is producing light-colored pheomelanin instead of dark-colored eumelanin. (Credit: [Djmirko](#); from [Wikimedia](#))

Melanocytes make two forms of melanin, eumelanin and pheomelanin. The relative amounts of these two pigments help determine the color of an individual's hair and skin. Individuals who produce mostly eumelanin tend to have brown or black hair and dark skin that tans easily (in humans). Eumelanin also protects skin from damage caused by ultraviolet (UV) radiation in sunlight. Individuals who produce mostly pheomelanin tend to have red or blond hair, freckles, and light-colored skin that tans poorly. Because pheomelanin does not protect skin from UV radiation, people with more pheomelanin have an increased risk of skin damage caused by sun exposure.

The melanocortin 1 receptor controls which type of melanin is produced by melanocytes. When the receptor is activated, it triggers a series of chemical reactions inside melanocytes that stimulate these cells to make eumelanin. If the receptor is not activated or is blocked, melanocytes make pheomelanin instead of eumelanin. This means that if the receptor is working correctly and is turned on, dark pigment will be produced. If the receptor is not functional or is not turned on, light pigment will be produced.



Figure 8.25 The three recognized colors of labs are due to black eumelanin, brown eumelanin, or pheomelanin. (Credit: [Erikeltic](#), from [Wikimedia](#))

Schmutz et. al. (2002) determined the DNA sequence for the MC1R gene from dogs of various colors. They determined that black and brown dogs all have one allele of MC1R, while yellow and red dogs have a different allele. The allele that leads to yellow or red color has a premature stop codon which results in a shorter-than-normal protein. This protein would be predicted to not function correctly. Remember that when the melanocortin 1 receptor is not functioning correctly, light pheomelanin pigment is produced and not dark eumelanin.

Dogs that are homozygous for the functioning allele of MC1R (which would cause eumelanin to be produced) are dark in color. Dogs that are homozygous for the non-functioning allele (which would cause pheomelanin to be produced) are light in color. Dogs that are heterozygous are dark in color. What does this tell you about which allele is dominant? If you said “the dark allele is dominant because it covers up the light allele”, you’re correct. We will use “E” to represent the genotype at MC1R because the dominant phenotype in this case is the production of eumelanin. Dogs that have the genotype EE or Ee will produce eumelanin and be dark. Dogs that have the genotype “ee” will produce pheomelanin and be light.

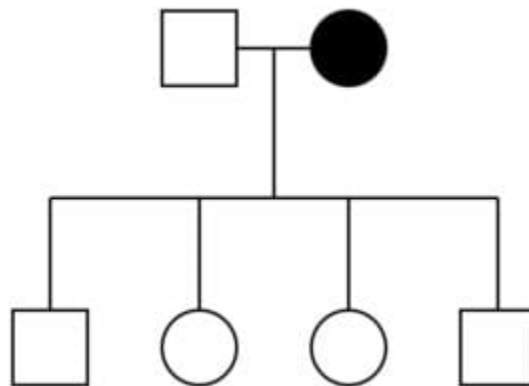


Figure 8.26 In this pedigree, the shaded individual is yellow. She therefore has the genotype ee and produces pheomelanin. We can’t tell the genotype of her mate by looking (he could be Ee or EE), but since all of their puppies were dark in color, we would predict that his genotype was EE. In this cross: EE x ee, 100% of the puppies would have the genotype Ee, so 100% of the puppies would produce eumelanin instead of pheomelanin.

The cross shown in Figure 8.26 can also be shown as a Punnett square (Figure 8.27). Since we are unsure whether the male dog has the genotype “EE” or “Ee”, we have to make two Punnett squares. Since all of the puppies resulting from this cross were black, we would predict that the first Punnett square shows the cross. However, it is possible that the second Punnett square is correct. There are only 4 puppies, so it’s not hard to imagine that they could all be black even though the Punnett square predicts only

50% black. It would be comparable to flipping a coin 4 times and getting 4 heads in a row. Getting 4 heads in a row is less likely, but definitely possible.

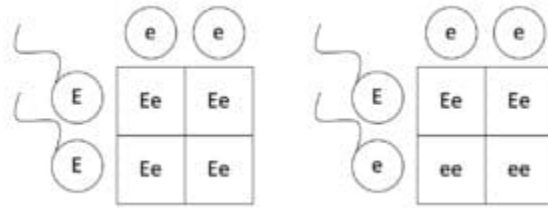


Figure 8.27 Cross from Figure 8.26 shown as Punnett squares.

It is very important to note here that yellow dogs still have the TYRP1 gene, even though they are not black or brown!

Human Connection

Common variations (polymorphisms) in the MC1R gene are associated with normal differences in skin and hair color. Certain genetic variations are most common in people with red hair, fair skin, freckles, and an increased sensitivity to sun exposure. These MC1R polymorphisms reduce the ability of the melanocortin 1 receptor to stimulate eumelanin production, causing melanocytes to make mostly pheomelanin. Although MC1R is a key gene in normal human pigmentation, researchers believe that the effects of other genes also contribute to a person's hair and skin coloring.

The melanocortin 1 receptor is also active in cells other than melanocytes, including cells involved in the body's immune and inflammatory responses. The receptor's function in these cells is unknown.



Photo credit: [dusdin on flickr](#); from [Wikipedia](#).

Resources

Unless otherwise noted, text and images by Lisa Bartee, 2016.

OpenStax, Biology. OpenStax CNX. May 27, 2016 <http://cnx.org/contents/s8Hh0oOc@9.10:FtsD6vMd@3/Mendels-Experiments>

OpenStax, Biology. OpenStax CNX. May 27, 2016 <http://cnx.org/contents/s8Hh0oOc@9.10:sbdXt0s3@4/Laws-of-Inheritance>

OpenStax, Biology. OpenStax CNX. May 27, 2016 <http://cnx.org/contents/s8Hh0oOc@9.10:zLLYW2hj@5/Extensions-of-the-Laws-of-Inheritance>

Human Connection – information about MC1R: “MC1R” by [Genetics Home Reference: Your Guide to Understanding Genetic Conditions](#), National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

Schmutz SM, Moker JS, Berryere TG, Christison KM, Dolf G. 2001. [An SNP is used to map MC1R to dog chromosome 5](#). Anim Genet. 32(1):43-4.

Schmutz SM, Berryere TG, Goldfinch AD. 2002. [TYRP1 and MC1R genotypes and their effects on coat color in dogs](#). Mammalian Genome 13, 380-387.

This page titled [3.1.5: Yellow fur color - a recessive trait](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Lisa Bartee, Walter Shriner and Catherine Creech.

3.1.6: Brindle color - partial dominance and epistasis

Brindle coloration is a black and brown striping pattern that is caused by different alleles at the “K locus”, which is probably a gene called ASIP that controls pigment switching (Figure 8.28); Ciampolini, 2013). There are three alleles of the K locus: K^B , k^{br} , and k^y (Kerns, 2007). The K^B allele is dominant over the other two alleles and produces solid black color. k^{br} produces the brindle color pattern and is dominant over the k^y allele. This means that dogs with the genotype $k^{br}k^{br}$ or $k^{br}k^y$ will have the brindle color pattern. Dogs with the genotype $k^y k^y$ are yellow in color.



Figure 8.28 This boxer shows the brindle color pattern, which looks sort of like tiger stripes. (Credit: Steve Henderson Location: Memphis, TN)

The K locus and MC1R (which controls the difference between dark eumelanin and light pheomelanin production) have an epistatic relationship. If a dog has two recessive alleles for MC1R and is therefore unable to make eumelanin, the dog will appear yellow regardless of its genotype at the K locus.

References

Unless otherwise noted, text and images by Lisa Bartee, 2016.

Ciampolini R, Cecchi F, Spaterna A, Bramante A, Bardet SM, Oulmouden A. 2013. [Characterization of different 5'-untranslated exons of the ASIP gene in black-and-tan Doberman Pinscher and brindle Boxer dogs](#). Anim Genet. 44(1):114-7.

Kerns JA, Cargill EJ, Clark LA, Candille SI, Berryere TG, Olivier M, Lust G, Todhunter RJ, Schmutz SM, Murphy KE, Barsh GS. 2007. [Linkage and segregation analysis of black and brindle coat color in domestic dogs](#). Genetics. 176(3):1679-89.

This page titled [3.1.6: Brindle color - partial dominance and epistasis](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Lisa Bartee, Walter Shriner and Catherine Creech.

3.1.7: White spotting - when there's more than two alleles

So far, we have discussed genes which have only two alleles. However, that is not always the case: there can be more than two alleles for a given gene. One example is the *MITF* gene, which is the major gene that controls white spotting in dogs. This protein is required for the migration and survival of melanocytes into the skin during development. If it is not functional, it impairs the ability of the skin to make pigment, thus “covering up” the effect of other color genes. There are thought to be at least four alleles that can contribute (Karlsson, 2007). Depending on which alleles are present in a dog, the amount of white can vary from none (a solid-colored dog) to mostly white (Table 8.1) and (Figure 8.29).

Table 8.1 Combinations of different alleles for *MITF* result in different amounts of white present in the coat.

Alleles	Amount of white
SS	None (solid colored)
Ss^i	Small amounts of white possible on face, chest, feet, and tail tip
Ss^p	More than 50% solid colored, with white on the face, chest, feet, collar, underbelly, and tail tip
$s^i s^p$	Approximately even amounts of color and white
$s^i s^e$	More than 50% white with irregular splashes of color
$s^e s^e$	Mostly white with only minimal areas of color, perhaps on one or both ears, an eye patch, or a spot near the tail



Figure 8.29 These dogs have different combinations of alleles of the *MITF* gene. The first dog probably has the genotype “ Ss^p ”; the dog in the center is likely “ SS ”; the dog on the right is likely “ $s^e s^e$ ”. (Credits: Funny black dog by X posid from [Publicdomainpictures](#). A black and white dog by Petr Kratochvil from Free stock photos. White dog with black ears by RetyiRetyi from Pixabay.)

Human Connection – Blood Type

Human blood type was discussed in the previous section. You may remember that there are three alleles for the ABO gene: A, B, and O. A and B are codominant, meaning that if both alleles are present, both will be seen in the phenotype. A person with type AB blood has one A allele and one B allele.

O is recessive to A and B. A person with the genotype AO will have Type A blood. A person with the genotype BO will have type B blood. Type O blood results from two O alleles.

	A	B	O
A	AA	AB	AO
B	AB	BB	BO
O	AO	BO	OO

Photo credit: [Kalaiaresy](#), from [Wikipedia](#).

References

Unless otherwise noted, text and images by Lisa Bartee, 2016.

Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ 3rd, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kämpe O, Hedhammar A, Lander ES, Andersson G, Andersson L, Lindblad-Toh K. 2007. [Efficient mapping of mendelian traits in dogs through genome-wide association](#). Nat Genet. 39(11):1321-8.

This page titled [3.1.7: White spotting - when there's more than two alleles](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Lisa Bartee, Walter Shriner and Catherine Creech.

3.1.8: KIT - embryonic lethality

Although MITF is the major gene impacting white spotting in dogs, a second gene known as KIT has also been shown to have an impact in a subpopulation of German Shepherd dogs. This mutation in KIT arose very recently (it appeared spontaneously in a female dog born in 2000) and causes a phenotype called “panda spotting.” The KIT gene produces a tyrosine kinase receptor protein that functions in the same melanogenesis pathway as MITF. The KIT receptor controls many important processes within the cells including growth and division, survival, and cell migration. Its signaling function is important in the development of many different types of cells, including melanocytes which produce the pigment melanin.

Interestingly, no dogs have been identified that are homozygous for the mutation in KIT. This is probably because having two mutations in this gene is lethal because of its many important functions. This is called a homozygous **lethal allele** that results in embryonic lethality. There are other examples of embryonic lethal genes:

- The Agouti gene causes coat color changes in many animals, including mice. One mutation that has been isolated in mice causes a yellow colored coat in mice heterozygous for the mutation. Homozygous mice die in the womb (Figure 8.30).

	A	A ^y
A	Agouti coat AA 	Yellow Coat AA ^y 
A ^y	Yellow coat AA ^y 	Dead A ^y A ^y 

Figure 8.30 Lethality of homozygous agouti mutation. Photo credit [Jcfidy](#); [Wikimedia](#).

- Mutations in the FGFR3 gene result in Achondroplasia, a form of dwarfism in humans (Figure 8.31). This disorder is inherited in an autosomal dominant fashion. Two dominant alleles results in lethality.



Figure 8.31 Jason Acuña outside of the Waterfront Marriott in Portland, OR. on August 15, 2009. Photo credit [Sakibomb222; Wikimedia](#).

- Manx cats have a dominant mutation that leads to them being tailless or having short, stubby tails (Figure 8.32). Kittens with two dominant alleles are thought to not survive until birth (Catworld, 2009).



Figure 8.32 A “rumpy riser” manx kitten. Photo credit Michelle Wiegold; [Wikimedia](#).

- Mexican hairless dogs contain a mutation in the *FOXI3* gene. Heterozygous dogs are hairless (Figure 8.33). Homozygous mutant dogs are never observed.



Figure 8.33 Michelob (Michael) and Irish Mist (Misty), a hairless and a coated Xoloitzcuintli. Christopher A. and Amanda L. Dellario, Nottingham, NH, USA. Photo taken by Amanda L. Dellario, August 2006. [Wikimedia](#).

In humans, the KIT gene is located on chromosome 4 (Figure 8.34), while it is located on chromosome 13 in dogs (Wong, 2012).

Cytogenetic Location: 4q12, which is the long (q) arm of [chromosome 4](#) at position 12

Molecular Location: base pairs 54,657,928 to 54,740,715 on chromosome 4 (Homo sapiens Annotation Release 108, GRCh38.p7) ([NCBI](#))

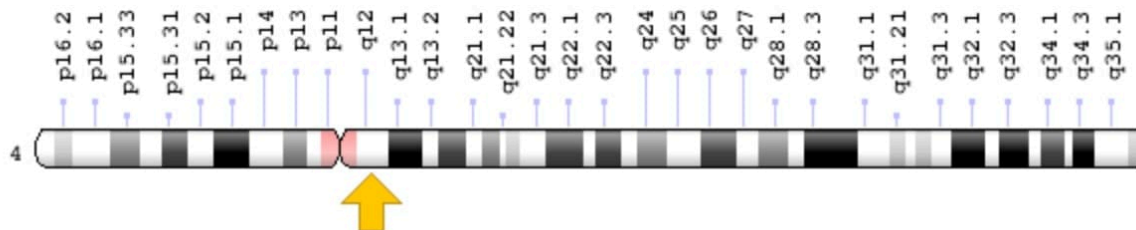


Figure 8.34 Chromosomal location of the KIT gene in humans. Photo credit [Genetics Home Reference](#); Public Domain.

While KIT has only been shown to have a phenotypic effect in one family of German Shepherd dogs (at least so far), at least 69 mutations in the human KIT gene have been identified (GHR, 2018). Mutations in the human KIT gene lead to piebaldism, where melanocytes are absent from certain areas of the hair and skin (Figure 8.35). These mutations are inherited in an autosomal dominant fashion. I couldn't find any information about whether homozygous piebald mutations are lethal in humans. This is likely because generating a homozygous piebald human would require mating two heterozygous piebald humans, and this disorder is rare so this mating would be unlikely.

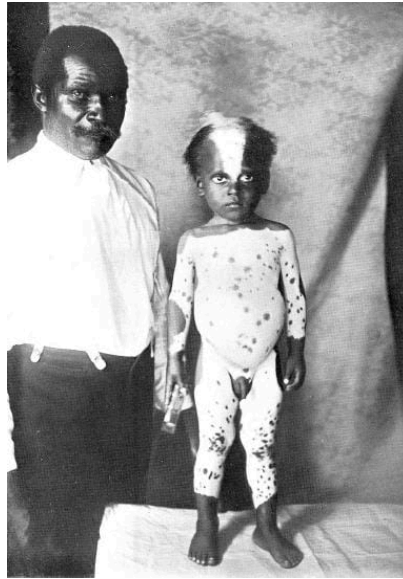


Figure 8.35 Piebaldism in a 5 year old boy. Neither of his parents nor any of his five siblings showed any white spotting. Photo credit [Wellcome Images; Wikimedia](#).

References

Wilson J. 2009. Cat World [Internet]. Feline Genetic Loci Table. [accessed on January 2, 2017]. Available from: <https://www.cat-world.com.au/feline-...oci-table.html>

Genetics Home Reference (GHR). 2018. National Institutes of Health. Available from <https://ghr.nlm.nih.gov/gene/KIT#conditions>

Wong AK, Ruhe AL, Roberson KR, Lowe ER, Williams DC, Neff MW. 2012. A *de novo* mutation in *KIT* causes white spotting in a subpopulation of German Shepherd dogs. *Animal Genetics*: 44:305-310.

3.1.9: Hemophilia - a sex-linked disorder

So far, all the genes we have discussed have had two copies present in all individuals. This is because the individual inherited one from the male parent's haploid gamete and one from the female parent's haploid gamete. The two gametes came together during fertilization to produce a diploid individual. There is, however, one exception to this: genes which are present on the sex chromosomes.

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**.

The X chromosome is one of two sex chromosomes. Humans and most mammals have two sex chromosomes, the X and Y. Females have two X chromosomes in their cells, while males have X and Y chromosomes in their cells. Egg cells all contain an X chromosome, while sperm cells contain an X or a Y chromosome. This arrangement means that during fertilization, it is the male that determines the sex of the offspring since the female can only give an X chromosome to the offspring.

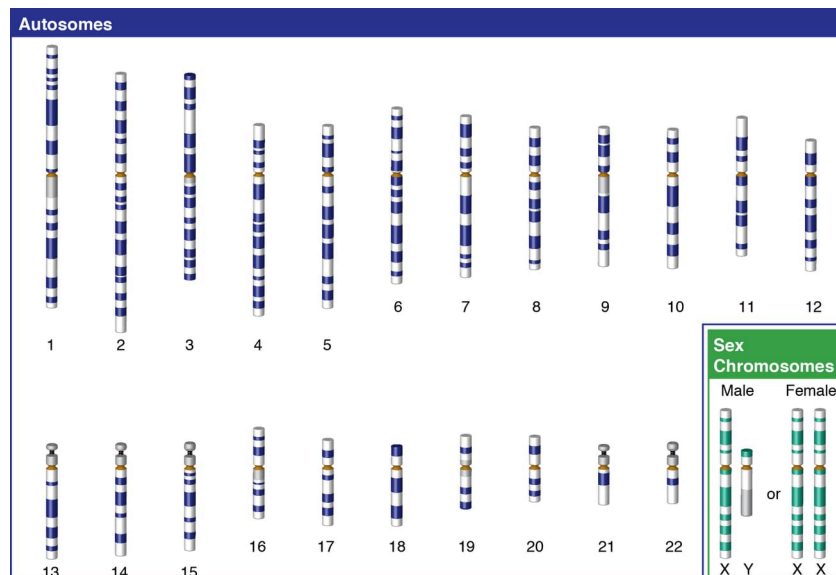


Figure 8.36 A diagram showing the autosomal and sex chromosomes. Remember that in a diploid cell, there would be two copies of each autosomal chromosome present. (Credit: Darryl Lega, NHGRI)

Most sex-linked genes are present on the X chromosome simply because it is much larger than the Y chromosome. The X chromosome spans about 155 million DNA base pairs and represents approximately 5 percent of the total DNA in cells. The X chromosome likely contains 800 to 900 genes. In contrast, the Y chromosome has approximately 59 million base pairs and only 50-60 genes. Sex is determined by the SRY gene, which is located on the Y chromosome and is responsible for the development of a fetus into a male. This means that the presence of a Y chromosome is what causes a fetus to develop as male. Other genes on the Y chromosome are important for male fertility.

Hemophilia is a bleeding disorder that slows the blood clotting process. People with this condition experience prolonged bleeding or oozing following an injury, surgery, or having a tooth pulled. In severe cases of hemophilia, continuous bleeding occurs after minor trauma or even in the absence of injury (spontaneous bleeding). Serious complications can result from bleeding into the joints, muscles, brain, or other internal organs. Milder forms of hemophilia do not necessarily involve spontaneous bleeding, and the condition may not become apparent until abnormal bleeding occurs following surgery or a serious injury.

The major types of this condition are hemophilia A (also known as classic hemophilia or factor VIII deficiency) and hemophilia B (also known as Christmas disease or factor IX deficiency). Although the two types have very similar signs and symptoms, they are

caused by mutations in different genes.

Hemophilia A and hemophilia B are inherited in an X-linked recessive pattern. The genes associated with these conditions are located on the X chromosome, which is one of the two sex chromosomes. In males (who have only one X chromosome), one altered copy of the gene in each cell is sufficient to cause the condition. In females (who have two X chromosomes), a mutation would have to occur in both copies of the gene to cause the disorder. Because it is unlikely that females will have two altered copies of this gene, it is very rare for females to have hemophilia. A characteristic of X-linked inheritance is that fathers cannot pass X-linked traits to their sons (Figure 8.37 and 8.38).

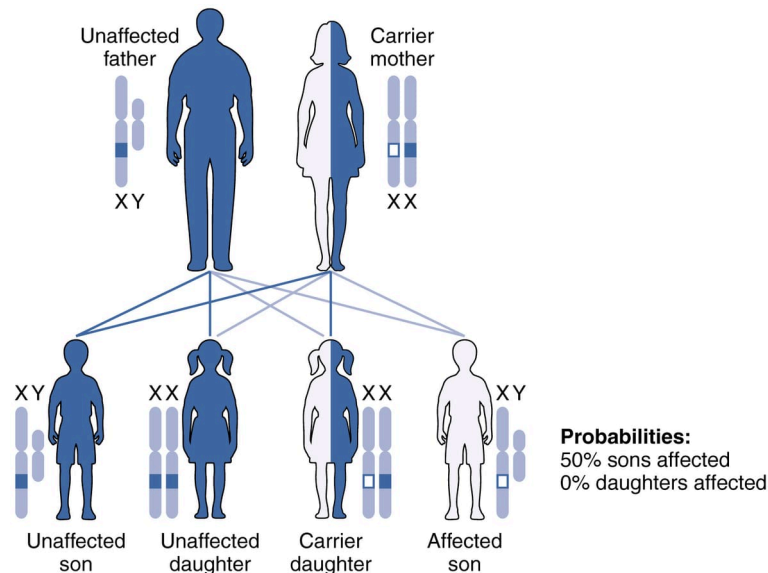


Figure 8.37 X-linked recessive inheritance. Photo credit OpenStax College; [OpenStax Anatomy and Physiology](#).

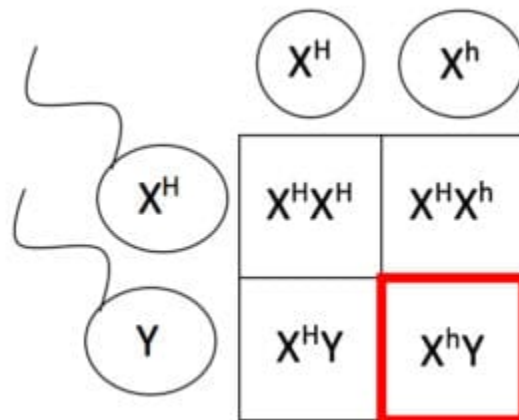


Figure 8.38 If a carrier female and a normal male produce offspring, there is a 25% total chance that they will have a child with hemophilia. None of their daughters will have the disease (although half will be carriers). Half their sons will be hemophiliacs.

In X-linked recessive inheritance, a female with one altered copy of the gene in each cell is called a carrier. Carrier females have about half the usual amount of coagulation factor VIII or coagulation factor IX, which is generally enough for normal blood clotting. However, about 10 percent of carrier females have less than half the normal amount of one of these coagulation factors; these individuals are at risk for abnormal bleeding, particularly after an injury, surgery, or tooth extraction.

Colorblindness is another example of a sex-linked trait in humans. The genes that produce the photopigments necessary for color vision are located on the X chromosome. If one of these genes is not functional because it contains a harmful mutation, the individual will be colorblind. Men are much more likely than women to be colorblind: up to 100 times more men than women have various types of colorblindness (<http://www.colour-blindness.com/general/prevalence/>).

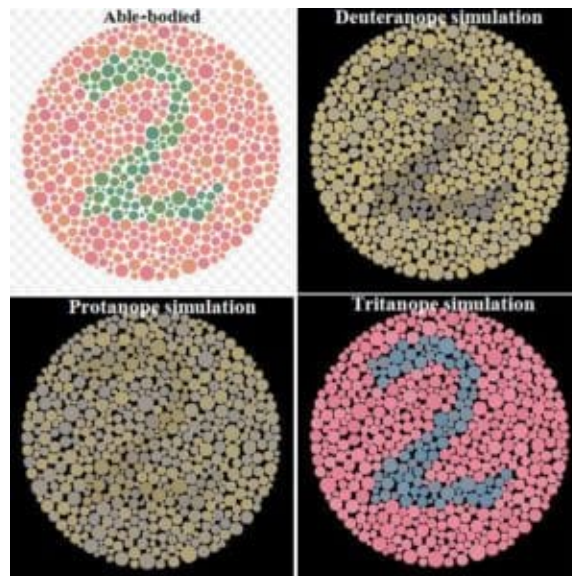


Figure 8.39 A test image for color-blindness as seen by someone with normal color vision and several types of colorblindness. (Credit: Sakurambo)

References:

Unless otherwise noted, text adapted by and images by Lisa Bartee, 2016.

OpenStax, Biology. OpenStax CNX. May 27, 2016 <http://cnx.org/contents/s8Hh0oOc@9.10:zLLYW2hj@5/Extensions-of-the-Laws-of-Inhe>

“X chromosome” by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

“Y chromosome” by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

“Hemophilia” by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

3.1.10: It's not all in the genes - the effect of environment

Not all traits are directly caused by DNA alone. The environment also plays a large role in shaping an individual's traits. Some examples can be seen below.

- **Height and weight:** A number of genes interact to determine the general height and weight that a person will have. But the environment has a major influence as well. If an individual is malnourished, their growth may be slowed and they may be smaller than they would have been if they had gotten enough food. In contrast, if a person consumes more calories than they need, their weight will likely increase regardless of their genetics.
- **Fingerprints:** the general characteristics of a person's fingerprints are determined by genetics, but the specific pattern is generated randomly during development. Identical twins typically have fingerprints that are similar, but not identical.
- **Intelligence:** Like most aspects of human behavior and cognition, intelligence is a complex trait that is influenced by both genetic and environmental factors. Roughly 50% of a person's IQ appears to be determined by genetic factors. Factors related to a child's home environment and parenting, education and availability of learning resources, and nutrition, among others, also contribute to intelligence. A person's environment and genes influence each other, and it can be challenging to tease apart the effects of the environment from those of genetics. For example, if a child's IQ is similar to that of his or her parents, is that similarity due to genetic factors passed down from parent to child, to shared environmental factors, or (most likely) to a combination of both? It is clear that both environmental and genetic factors play a part in determining intelligence.
- **Cancer Risk:** For example, a person could inherit a mutation in the BRCA1 gene, which increases the risk of developing breast or ovarian cancer. Researchers have identified more than 1,800 mutations in the BRCA1 gene. Most BRCA1 gene mutations lead to the production of an abnormally short version of the BRCA1 protein or prevent any protein from being made from one copy of the gene. As a result, less of this protein is available to help repair damaged DNA or fix mutations that occur in other genes. As these defects accumulate, they can trigger cells to grow and divide uncontrollably to form a tumor. These mutations are present in every cell in the body and can be passed from one generation to the next. As a result, they are associated with cancers that cluster in families. However, not everyone who inherits a mutation in the BRCA1 gene will develop cancer. Other genetic, environmental, and lifestyle factors also contribute to a person's cancer risk.
- In contrast, cancer can be caused by purely environmental factors. According to the [CDC](#), cigarette smoking is the number one risk factor for lung cancer. In the United States, cigarette smoking is linked to about 90% of lung cancers and people who smoke are 15 to 30 times more likely to get lung cancer or die from lung cancer than people who do not smoke. Radon exposure also increases the likelihood that a person will develop lung cancer.



Figure 8.40 The colors on the poodle seen in this figure have no relationship to his DNA: he was dyed for a parade. (Credit: [skeeze](#))

References

“BRCA1” by [Genetics Home Reference: Your Guide to Understanding Genetic Conditions](#), National Institutes of Health: U.S.

National Library of Medicine is in the Public Domain

“Is intelligence determined by genetics?” by Genetics Home Reference: Your Guide to Understanding Genetic Conditions, National Institutes of Health: U.S. National Library of Medicine is in the Public Domain

This page titled [3.1.10: It's not all in the genes - the effect of environment](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Lisa Bartee, Walter Shriner and Catherine Creech.

3.1.11: Pleiotropy - one gene affects more than one trait

So far, we have discussed examples where changing the DNA sequence of one gene affects one protein, which affects one specific trait (for example, the change from brown to black fur). However, there are examples where one mutation can affect more than one trait. This is called **pleiotropy**.

MC1R

MC1R, the gene which leads to the difference between yellow and dark colored dogs, is also found in humans. Recall that MC1R helps determine whether mostly eumelanin (dark pigment) or pheomelanin (lighter reddish pigment) will be produced. Humans who produce mostly eumelanin have the active allele of MC1R and have darker skin that tans easily and darker hair. Humans who produce mostly pheomelanin have inactive MC1R proteins and have red or blonde hair and light skin with freckles that does not tan easily.

One relatively obvious secondary phenotype of having lighter skin is that rates of skin cancer are higher in those individuals compared to individuals who have darker skin that tans easily. Since MC1R affects skin pigmentation, it also has an effect on skin cancer rates. In addition, MC1R has an effect on cancer rates that is unrelated to skin pigmentation due to its interactions with other genes that regulate inflammatory responses, DNA repair, and apoptosis (Feller, 2016).

Interestingly, red headed individuals also exhibit higher pain tolerance due to their MC1R alleles (Liem, 2004; Liem, 2006). So far, the reason for this is unknown.

Mutations in MC1R have been shown to decrease knee cartilage in mice (Lorenz, 2014). Again, the mechanism for this is not understood, but likely relates to MC1Rs signaling role.

MC1R polymorphisms have been associated with a decrease in the development of sepsis (blood poisoning) in humans (Seaton, 2017). This is probably due to the role that MC1R plays in inflammation (again, due to its role in signal transduction).

Activation of MC1R by agonists (chemicals that activate receptors) were shown to reduce several harmful symptoms in the kidneys of rats (Lindskog, 2010). The specific reason for this is not understood.

“Frizzled” chickens

The dominant “frizzle” allele causes feathers to turn upwards rather than remain flat against the chicken’s body (Figure 3.1.11.1). However, along with producing defective feathers, the frizzle allele also leads to abnormal body temperatures, higher metabolic and blood flow rates, and greater digestive capacity. Furthermore, chickens who have this allele also lay fewer eggs than their wild-type counterparts.



Figure 8.41 A “frizzled” chicken. Photo credit Joe Goldberg; [Flickr](#).

Phenylketonuria

Phenylketonuria (PKU) is a disorder that affects the levels of the amino acid phenylalanine in the body. We get phenylalanine from food, then process it within our cells. Individuals with PKU have a mutation in the enzyme required to break down phenylalanine. Levels of this amino acid build up in individuals with PKU. This build up can lead to a variety of different symptoms including intellectual disability, seizures, poor bone strength, skin rashes, behavioral and mental disorders, and an unusually small head. If you’ve ever seen a warning on a package that says “Phenylketonurics – contains phenylalanine”, this is why.



References

- Feller L, Khammissa RAG, Kramer B, Altini M, Lemmer J. 2016. Basal cell carcinoma, squamous cell carcinoma and melanoma of the head and face. *Head Face Med.* 12:11.
- Liem EB, et. al. 2004. Anesthetic Requirement is Increased in Redheads. *Anesthesiology.* 101(2): 279-283.
- Liem EB, Joiner TV, Tsueda K, Sessler DI. 2006. Increased sensitivity to Thermal Pain and Reduced Subcutaneous Lidocaine Efficacy in Redheads. *Anesthesiology.* 102(3): 509-514.
- Lindskog A, et. al. 2010. Melanocortin 1 Receptor Agonists Reduce Proteinuria. *J Am Soc Nephron.* 21(8): 1290-1298.
- Lorenz J, et. al. 2014. Melanocortin 1 Receptor-Signaling Deficiency Results in an Articular Cartilage Phenotype and Accelerates Pathogenesis of Surgically Induced Murine Osteoarthritis. *PLoS One.* 9(9): e105858.
- Seaton ME, Parent BA, Mood RF, Wurfel MM, Muffley LA, O’Keefe GE, Gibran NS. 2017. MELANOCORTIN-1 RECEPTOR POLYMORPHISMS AND THE RISK OF COMPLICATED SEPSIS AFTER TRAUMA: A CANDIDATE GENE ASSOCIATION STUDY. *Shock.* 47(1): 79-85.

This page titled [3.1.11: Pleiotropy - one gene affects more than one trait](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by Lisa Bartee, Walter Shriner and Catherine Creech.

4: GENETIC TECHNOLOGY

4.1: Biotechnology

4.2: Mapping Genomes

4.3: Whole-Genome Sequencing

4.4: Genomics and Proteomics

4.4.1: Biotechnology

4.1: Biotechnology

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 9.1). 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.

DNA Extraction

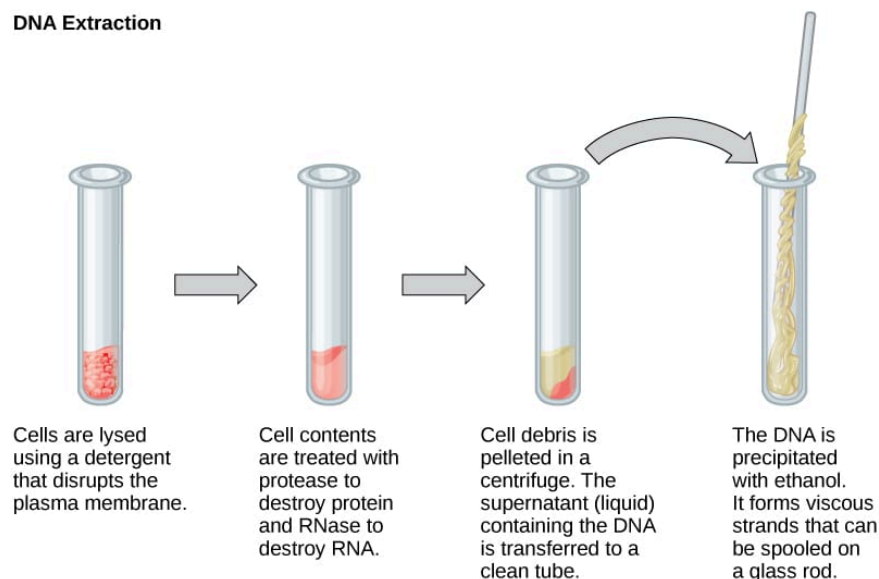


Figure 9.1 This diagram shows the basic method used for the extraction of DNA.

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 9.2). 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.

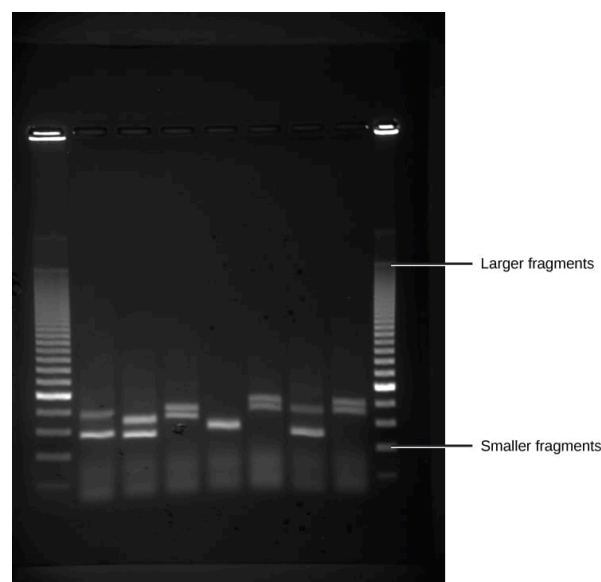


Figure 9.2 Shown are DNA fragments from six samples run on a gel, stained with a fluorescent dye and viewed under UV light. (credit: modification of work by James Jacob, Tompkins Cortland Community College)

Polymerase Chain Reaction

Downstream molecular DNA analysis frequently requires researchers to isolate and focus on highly specific regions of a genome. However, scientists routinely encounter experimental scenarios where only a few fragments of a DNA molecule are available for study. Because these small quantities are entirely insufficient for standard laboratory visualization procedures, such as gel electrophoresis, scientists rely on the **polymerase chain reaction (PCR)**. PCR is a fundamental biotechnology technique used to rapidly amplify and increase the copy number of targeted regions of DNA for further analyses.

To achieve exponential amplification, the reaction uses a specialized, heat-stable form of DNA polymerase, called **Taq Polymerase**, alongside short, synthetic nucleotide sequences called **primers**. These primers are custom-designed to bind via complementary base pairing to the exact boundaries of the DNA segment targeted for replication. The reaction mixture contains the target template DNA, a pair of forward and reverse primers, free-floating deoxynucleotide triphosphates (dNTPs) to serve as building blocks, *Taq* polymerase, and a chemically balanced buffer enriched with magnesium ions (Mg^{2+}) to stabilize the enzyme's activity.

To successfully amplify the target sequence, the polymerase chain reaction relies on a thermal cycler to repeat a three-step process—denaturation, annealing, and extension—over dozens of consecutive cycles. First, during the denaturation phase, the reaction mixture is heated to a high temperature, typically around 95°C. This heat breaks the hydrogen bonds holding the

complementary bases together, causing the double-stranded DNA template to separate into two isolated single strands. Second, during the annealing phase, the temperature is lowered to between 50°C and 65°C. This cooling allows the short, synthetic DNA primers to find and bind to their complementary target sequences on the now-separated single strands of template DNA. Finally, during the extension phase, the temperature is raised to an optimal level, usually around 72°C. At this temperature *Taq* polymerase binds to the primed single strands and begins incorporating free-floating nucleotides. The enzyme synthesizes a brand-new complementary strand by moving down the template, effectively duplicating the target region. Because each completed cycle doubles the amount of target DNA, repeating this three-step sequence thirty to forty times results in the exponential creation of billions of identical DNA copies in just a few hours.

Because of its sensitivity and efficiency, PCR has become indispensable across laboratory applications. It is widely used in forensic science to identify individuals from trace biological samples left at crime scenes, in genetic testing for paternity analysis, in evolutionary biology to compare fragile strands of ancient DNA with modern organisms, and in clinical medicine to amplify target genes prior to determining their exact nucleotide sequence.

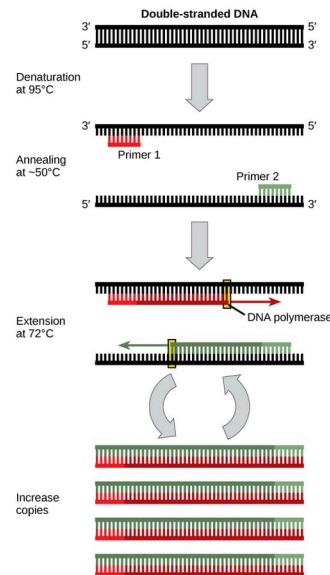


Figure 9.3 Polymerase chain reaction, or PCR, is used to produce many copies of a specific sequence of DNA using a special form of DNA polymerase.

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 9.4).

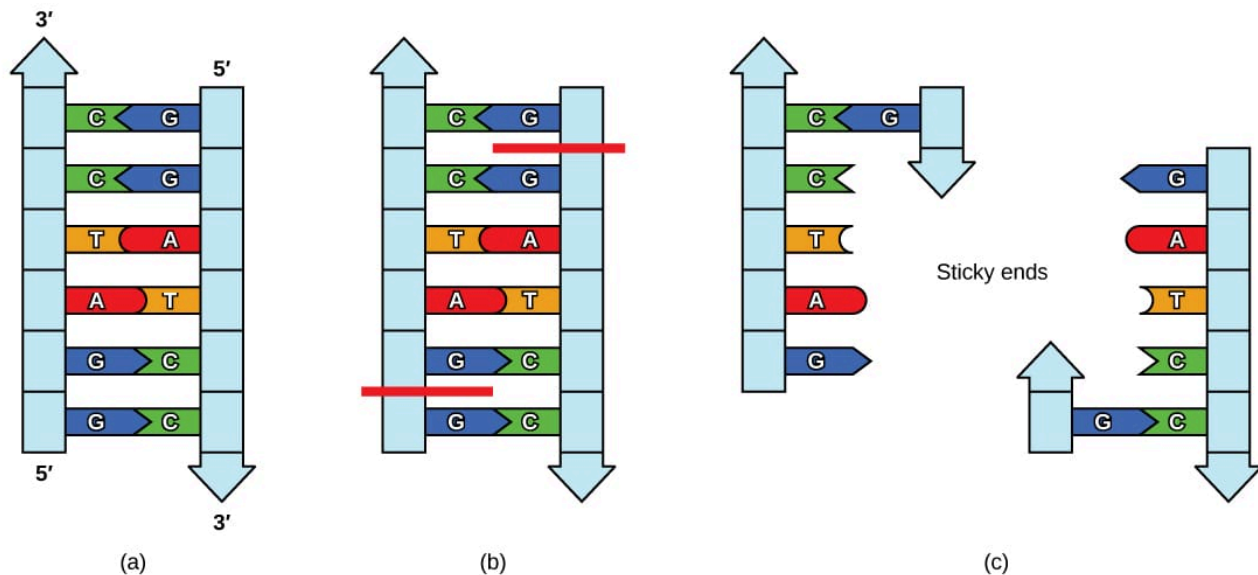


Figure 9.4 In this (a) six-nucleotide restriction enzyme recognition site, notice that the sequence of six nucleotides reads the same in the 5' to 3' direction on one strand as it does in the 5' to 3' direction on the complementary strand. This is known as a palindrome. (b) The restriction enzyme makes 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 9.5).

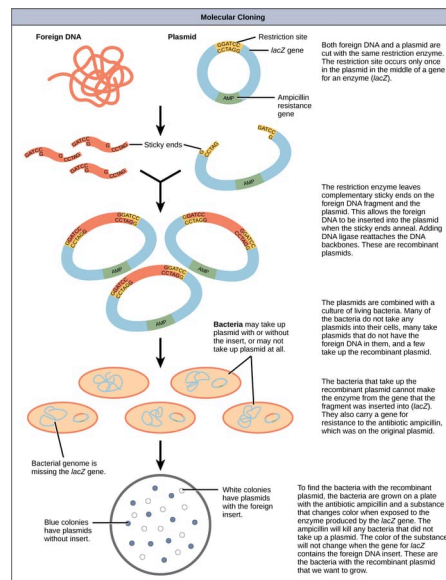


Figure 9.5 This diagram shows the steps involved in molecular cloning.

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.

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.

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.

Key Terms

Biotechnology

The use of living organisms, cells, or biological molecules to develop products and technologies for medicine, agriculture, industry, and research.

DNA Extraction

The process of isolating DNA from cells by breaking open the cells, removing proteins and other molecules, and purifying the DNA.

RNA Extraction

The process of isolating RNA while protecting it from degradation by RNases.

Gel Electrophoresis

A laboratory technique that separates DNA, RNA, or proteins based on their size and electrical charge using an electric field.

Polymerase Chain Reaction (PCR)

A technique used to amplify a specific region of DNA, producing millions to billions of copies.

Thermal Cycler

An instrument that rapidly changes temperatures during PCR to allow DNA denaturation, primer annealing, and DNA synthesis.

Taq Polymerase

A heat-stable DNA polymerase isolated from *Thermus aquaticus* that synthesizes DNA during PCR.

Primer

A short DNA sequence that binds to a target DNA region and provides a starting point for DNA polymerase.

Denaturation

The first step of PCR in which double-stranded DNA is heated to separate into single strands.

Annealing

The PCR step during which primers bind to complementary DNA sequences.

Extension

The PCR step during which DNA polymerase synthesizes a new DNA strand.

Molecular Cloning

The process of producing many identical copies of a specific DNA fragment by inserting it into a cloning vector.

Plasmid

A small circular DNA molecule found in bacteria that replicates independently of the bacterial chromosome.

Vector

A DNA molecule used to carry foreign DNA into a host cell.

Foreign DNA

DNA originating from another organism that is introduced into a host cell.

Restriction Enzyme (Restriction Endonuclease)

An enzyme that recognizes specific DNA sequences and cuts DNA at or near those sites.

Restriction Site

A specific DNA sequence recognized by a restriction enzyme.

Palindrome

A DNA sequence that reads the same in the 5' → 3' direction on both complementary strands.

Sticky Ends

Single-stranded DNA overhangs produced by staggered cuts made by restriction enzymes that facilitate joining DNA fragments.

Recombinant DNA

DNA molecules created by combining genetic material from two or more different sources.

Genetic Engineering

The deliberate modification of an organism's DNA using biotechnology.

Genetically Modified Organism (GMO)

An organism whose DNA has been altered using genetic engineering techniques.

Transgenic Organism

A genetically modified organism that contains DNA from a different species.

Check Your Understanding

1. What is biotechnology?
2. Why must DNA or RNA first be extracted before molecular analysis?
3. Why is RNA generally more difficult to isolate than DNA?
4. What is the purpose of detergents and enzymes during DNA extraction?
5. What is the difference between a genome and a single gene?
6. Explain how gel electrophoresis separates DNA fragments.
7. Why do smaller DNA fragments migrate farther through a gel than larger fragments?
8. Why is DNA able to move through an electric field?
9. What is the purpose of PCR?
10. What are the three major steps of a PCR cycle?
11. What happens during the denaturation step?
12. What role do primers play during PCR?
13. Why is Taq polymerase used instead of most other DNA polymerases?
14. Why does PCR produce an exponential increase in DNA?
15. What is molecular cloning?
16. Why are plasmids useful cloning vectors?
17. What is the function of restriction enzymes?
18. Why are sticky ends useful when constructing recombinant DNA molecules?
19. Explain why bacteria are commonly used to produce recombinant DNA and proteins.
20. What is genetic engineering?
21. What is the difference between a genetically modified organism (GMO) and a transgenic organism?

Contributors and Attributions

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at <https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd>).

This page titled [4.1: Biotechnology](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [10.1: Cloning and Genetic Engineering](#) by [OpenStax](#) is licensed [CC BY 4.0](#).

4.2: Mapping Genomes

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. Genome mapping is the process of finding the locations of genes on each chromosome. The maps that genome mapping create 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 an interstate highway map) and use genetic markers (similar to landmarks). A genetic marker is a gene or sequence on a chromosome that co-segregates (shows genetic linkage) with a specific trait. Early geneticists called this linkage analysis. Physical maps present the intimate details of smaller chromosome regions (similar to a detailed road map). 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 genome's complete picture. Having a complete genome 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 like cancer, heart disease, and cystic fibrosis. We can use genome mapping in a variety of other applications, such as using live microbes to clean up pollutants or even prevent pollution. Research involving plant genome mapping may lead to producing higher crop yields or developing plants that better adapt to climate change.

Genetic Maps

The study of genetic maps begins with linkage analysis, a procedure that analyzes the recombination frequency between genes to determine if they are linked or show independent assortment. Scientists used the term *linkage* before the discovery of DNA. Early geneticists relied on observing phenotypic changes to understand an organism's genotype. Shortly after Gregor Mendel (the father of modern genetics) proposed that traits were determined by what we now call genes, other researchers observed that different traits were often inherited together, and thereby deduced that the genes were physically linked by their location on the same chromosome. Gene mapping relative to each other based on linkage analysis led to developing the first genetic maps.

Observations that certain traits were always linked and certain others were not linked came from studying the offspring of crosses between parents with different traits. For example, in garden pea experiments, researchers discovered, that the flower's color and plant pollen's shape were linked traits, and therefore the genes encoding these traits were in close proximity on the same chromosome. We call exchanging DNA between homologous chromosome pairs genetic recombination, which occurs by crossing over DNA between homologous DNA strands, such as nonsister chromatids. Linkage analysis involves studying the recombination frequency between any two genes. The greater the distance between two genes, the higher the chance that a recombination event will occur between them, and the higher the recombination frequency between them. Figure 9.10 shows two possibilities for recombination between two nonsister chromatids during meiosis. If the recombination frequency between two genes is less than 50 percent, they are linked.

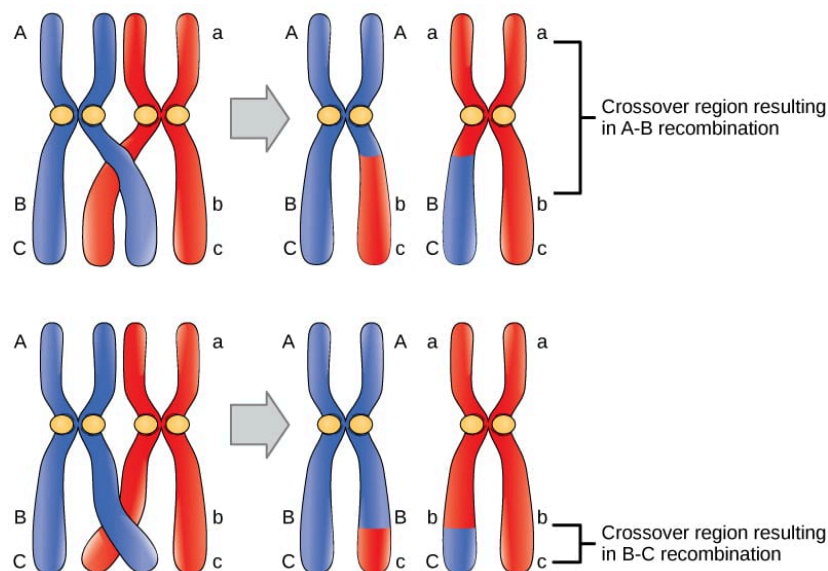


Figure 9.10 Crossover may occur at different locations on the chromosome. Recombination between genes *A* and *B* is more frequent than recombination between genes *B* and *C* because genes *A* and *B* are farther apart. Therefore, a crossover is more likely to occur between them.

The generation of genetic maps requires markers, just as a road map requires landmarks (such as rivers and mountains). Scientists based early genetic maps on using known genes as markers. Scientists now use more sophisticated markers, including those based on non-coding DNA, to compare individuals' genomes in a population. Although individuals of a given species are genetically similar, they are not identical. Every individual has a unique set of traits. These minor differences in the genome between individuals in a population are useful for genetic mapping purposes. In general, a good genetic marker is a region on the chromosome that shows variability or polymorphism (multiple forms) in the population.

Some genetic markers that scientists use in generating genetic maps are restriction fragment length polymorphisms (RFLP), variable number of tandem repeats (VNTRs), microsatellite polymorphisms, and the single nucleotide polymorphisms (SNPs). We can detect RFLPs (sometimes pronounced "rif-lips") when the DNA of an individual is cut with a restriction endonuclease that recognizes specific sequences in the DNA to generate a series of DNA fragments, which we can then analyze using gel electrophoresis. Every individual's DNA will give rise to a unique pattern of bands when cut with a particular set of restriction endonucleases. Scientists sometimes refer to this as an individual's **DNA fingerprint**. Certain chromosome regions that are subject to polymorphism will lead to generating the unique banding pattern. VNTRs are repeated sets of nucleotides present in DNA's non-coding regions. Non-coding, or junk, DNA has no known biological function; however, research shows that much of this DNA is actually transcribed. While its function is uncertain, it is certainly active, and it may be involved in regulating coding genes. The number of repeats may vary in a population's individual organisms. Microsatellite polymorphisms are similar to VNTRs, but the repeat unit is very small. SNPs are variations in a single nucleotide.

Because genetic maps rely completely on the natural process of recombination, natural increases or decreases in the recombination level given genome area affects mapping. Some parts of the genome are recombination hotspots; whereas, others do not show a propensity for recombination. For this reason, it is important to look at mapping information developed by multiple methods.

Physical Maps

A physical map provides detail of the actual physical distance between genetic markers, as well as the number of nucleotides. There are three methods scientists use to create a physical map: cytogenetic mapping, radiation hybrid mapping, and sequence mapping. Cytogenetic mapping uses information from microscopic analysis of stained chromosome sections (Figure 9.11). It is possible to determine the approximate distance between genetic markers using cytogenetic mapping, but not the exact distance (number of base pairs). Radiation hybrid mapping uses radiation, such as x-rays, to break the DNA into fragments. We can adjust the radiation amount to create smaller or larger fragments. This technique overcomes the limitation of genetic mapping, and we can adjust the radiation so that increased or decreased recombination frequency does not affect it. Sequence mapping resulted from DNA sequencing technology that allowed for creating detailed physical maps with distances measured in terms of the number of base pairs. Creating genomic libraries and complementary DNA (cDNA) libraries (collections of cloned sequences or all DNA from a genome) has sped the physical mapping process. A genetic site that scientists use to generate a physical map with sequencing technology (a sequence-tagged site, or STS) is a unique sequence in the genome with a known exact chromosomal location. An expressed sequence tag (EST) and a single sequence length polymorphism (SSLP) are common STSs. An EST is a short STS that we can identify with cDNA libraries, while we obtain SSLPs from known genetic markers, which provide a link between genetic and physical maps.

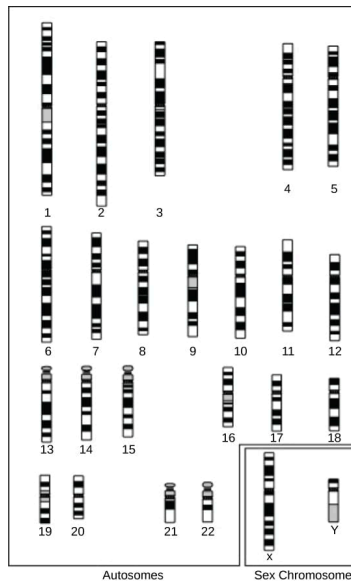


Figure 9.11 A cytogenetic map shows the appearance of a chromosome after scientists stain and exam it under a microscope. (credit: National Human Genome Research Institute)

Genetic and Physical Maps Integration

Genetic maps provide the outline and physical maps provide the details. It is easy to understand why both genome mapping technique types are important to show the big picture. Scientists use information from each technique in combination to study the genome. Scientists are using genomic mapping with different model organisms for research. Genome mapping is still an ongoing process, and as researchers develop more advanced techniques, they expect more breakthroughs. Genome mapping is similar to completing a complicated puzzle using every piece of available data. Mapping information generated in laboratories all over the world goes into central databases, such as GenBank at the National Center for Biotechnology Information (NCBI). Researchers are making efforts for the information to be more easily accessible to other researchers and the general public. Just as we use global positioning systems instead of paper maps to navigate through roadways, NCBI has created a genome viewer tool to simplify the data-mining process.

4.3: Whole-Genome Sequencing

Although there have been significant advances in the medical sciences in recent years, doctors are still confounded by some diseases, and they are using whole-genome sequencing to discover the root of the problem. Whole-genome sequencing is a process that determines an entire genome's DNA sequence. 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.

For example, whole-exome sequencing is a lower-cost alternative to whole genome sequencing. In exome sequencing, the doctor sequences only the DNA's coding, exon-producing regions. In 2010, doctors used whole-exome sequencing to save a young boy whose intestines had multiple mysterious abscesses. The child had several colon operations with no relief. Finally, they performed whole-exome sequencing, which revealed a defect in a pathway that controls apoptosis (programmed cell death). The doctors used a bone-marrow transplant to overcome this genetic disorder, leading to a cure for the boy. He was the first person to receive successful treatment based on a whole-exome sequencing diagnosis. Today, human genome sequencing is more readily available and results are available within two days for about \$1000.

Strategies Used in Sequencing Projects

The basic sequencing technique used in all modern day sequencing projects is the chain termination method (also known as the dideoxy method), which Fred Sanger developed in the 1970s. The chain termination method involves DNA replication of a single-stranded template by using a primer and a regular deoxynucleotide (dNTP), which is a monomer, or a single DNA unit. The primer and dNTP mix with a small proportion of fluorescently labeled dideoxynucleotides (ddNTPs). The ddNTPs are monomers that are missing a hydroxyl group (–OH) at the site at which another nucleotide usually attaches to form a chain (Figure 17.13). Scientists label each ddNTP with a different color of fluorophore. Every time a ddNTP incorporates in the growing complementary strand, it terminates the DNA replication process, which results in multiple short strands of replicated DNA that each terminate at a different point during replication. When gel electrophoresis processes the reaction mixture after separating into single strands, the multiple newly replicated DNA strands form a ladder because of the differing sizes. Because the ddNTPs are fluorescently labeled, each band on the gel reflects the DNA strand's size and the ddNTP that terminated the reaction. The different colors of the fluorophore-labeled ddNTPs help identify the ddNTP incorporated at that position. Reading the gel on the basis of each band's color on the ladder produces the template strand's sequence (Figure 9.13).

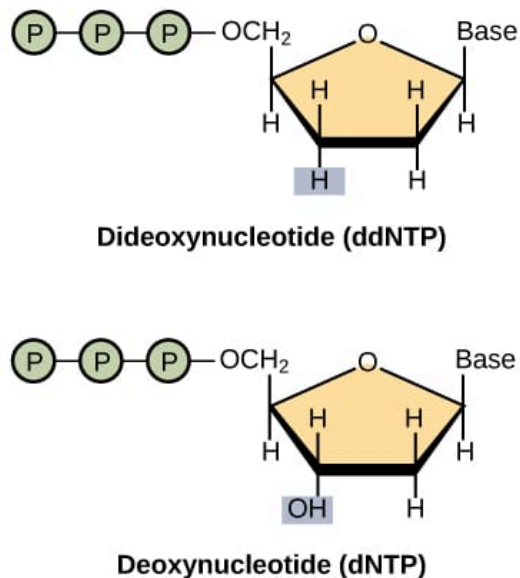


Figure 9.12 A dideoxynucleotide is similar in structure to a deoxynucleotide, but is missing the 3' hydroxyl group (indicated by the box). When a dideoxynucleotide is incorporated into a DNA strand, DNA synthesis stops.

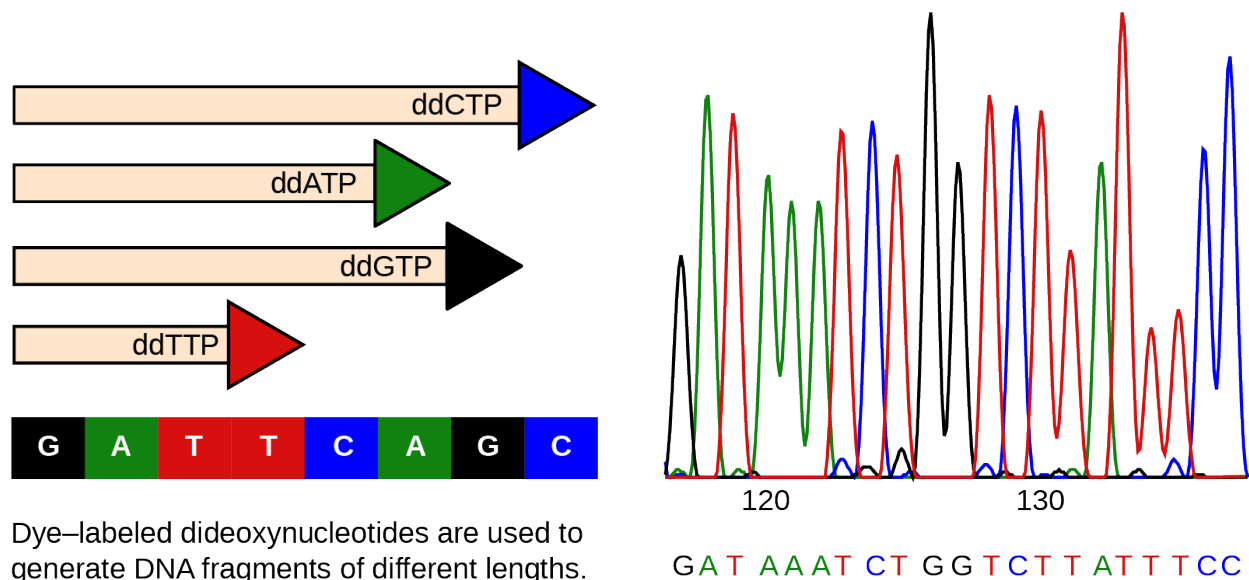


Figure 9.13 This figure illustrates Frederick Sanger's dideoxy chain termination method. Using dideoxynucleotides, the DNA fragment can terminate at different points. The DNA separates on the basis of size, and we can read these bands based on the fragments' size.

Early Strategies: Shotgun Sequencing and Pair-Wise End Sequencing

In shotgun sequencing method, several DNA fragment copies cut randomly into many smaller pieces (somewhat like what happens to a round shot cartridge when fired from a shotgun). All of the segments sequence using the chain-sequencing method. Then, with sequence computer assistance, scientists can analyze the fragments to see where their sequences overlap. By matching overlapping sequences at each fragment's end, scientists can reform the entire DNA sequence. A larger sequence that is assembled from overlapping shorter sequences is called a contig. As an analogy, consider that someone has four copies of a landscape photograph that you have never seen before and know nothing about how it should appear. The person then rips up each photograph with their hands, so that different size pieces are present from each copy. The person then mixes all of the pieces together and asks you to reconstruct the photograph. In one of the smaller pieces you see a mountain. In a larger piece, you see that the same mountain is behind a lake. A third fragment shows only the lake, but it reveals that there is a cabin on the shore of the lake. Therefore, from looking at the overlapping information in these three fragments, you know that the picture contains a mountain behind a lake that has a cabin on its shore. This is the principle behind reconstructing entire DNA sequences using shotgun sequencing.

Originally, shotgun sequencing only analyzed one end of each fragment for overlaps. This was sufficient for sequencing small genomes. However, the desire to sequence larger genomes, such as that of a human, led to developing double-barrel shotgun sequencing, or pairwise-end sequencing. In pairwise-end sequencing, scientists analyze each fragment's end for overlap. Pairwise-end sequencing is, therefore, more cumbersome than shotgun sequencing, but it is easier to reconstruct the sequence because there is more available information.

Next-generation Sequencing

Since 2005, automated sequencing techniques used by laboratories are under the umbrella of next-generation sequencing (deep sequencing or massively parallel sequencing), which is a group of automated techniques used for rapid DNA sequencing. These automated low-cost sequencers can generate sequences of hundreds of thousands or millions of short fragments (25 to 500 base pairs) in the span of one day. These sequencers use sophisticated software to get through the cumbersome process of putting all the fragments in order.

Comparing Sequences

A sequence alignment is an arrangement of proteins, DNA, or RNA. Scientists use it to identify similar regions between cell types or species, which may indicate function or structure conservation. We can use sequence alignments to construct phylogenetic trees. The following website uses a software program called [BLAST \(basic local alignment search tool\)](#).

Under “Basic Blast,” click “Nucleotide Blast.” Input the following sequence into the large “query sequence” box: ATTGCTTCGATTGCA. Below the box, locate the “Species” field and type “human” or “Homo sapiens”. Then click “BLAST” to compare the inputted sequence against the human genome’s known sequences. The result is that this sequence occurs in over a hundred places in the human genome. Scroll down below the graphic with the horizontal bars and you will see a short description of each of the matching hits. Pick one of the hits near the top of the list and click on “Graphics”. This will bring you to a page that shows the sequence’s location within the entire human genome. You can move the slider that looks like a green flag back and forth to view the sequences immediately around the selected gene. You can then return to your selected sequence by clicking the “ATG” button.

Use of Whole-Genome Sequences of Model Organisms

British biochemist and Nobel Prize winner Fred Sanger used a bacterial virus, the bacteriophage *φx174* (5368 base pairs), to completely sequence the first genome. Other scientists later sequenced several other organelle and viral genomes. American biotechnologist, biochemist, geneticist, and businessman Craig Venter sequenced the bacterium *Haemophilus influenzae* in the 1980s. Approximately 74 different laboratories collaborated on sequencing the genome of the yeast *Saccharomyces cerevisiae*, which began in 1989 and was completed in 1996, because it was 60 times bigger than any other genome sequencing. By 1997, the genome sequences of two important model organisms were available: the bacterium *Escherichia coli* K12 and the yeast *Saccharomyces cerevisiae*. We now know the genomes of other model organisms, such as the mouse *Mus musculus*, the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, and humans *Homo sapiens*. Researchers perform extensive basic research in model organisms because they can apply the information to genetically similar organisms. A model organism is a species that researchers use as a model to understand the biological processes in other species that the model organism represents. Having entire genomes sequenced helps with the research efforts in these model organisms. The process of attaching biological information to gene sequences is genome annotation. Annotating gene sequences helps with basic experiments in molecular biology, such as designing PCR primers and RNA targets.

Genome Sequence Uses

DNA microarrays are methods that scientists use to detect gene expression by analyzing different DNA fragments that are fixed to a glass slide or a silicon chip to identify active genes and sequences. We can discover almost one million genotypic abnormalities using microarrays; whereas, whole-genome sequencing can provide information about all six billion base pairs in the human genome. Although studying genome sequencing medical applications is interesting, this discipline dwells on abnormal gene function. Knowing about the entire genome will allow researchers to discover future onset diseases and other genetic disorders early. This will allow for more informed decisions about lifestyle, medication, and having children. Genomics is still in its infancy, although someday it may become routine to use whole-genome sequencing to screen every newborn to detect genetic abnormalities.

In addition to disease and medicine, genomics can contribute to developing novel enzymes that convert biomass to biofuel, which results in higher crop and fuel production, and lower consumer cost. This knowledge should allow better methods of control over the microbes that industry uses to produce biofuels. Genomics could also improve monitoring methods that measure the impact of pollutants on ecosystems and help clean up environmental contaminants. Genomics has aided in developing agrochemicals and pharmaceuticals that could benefit medical science and agriculture.

It sounds great to have all the knowledge we can get from whole-genome sequencing; however, humans have a responsibility to use this knowledge wisely. Otherwise, it could be easy to misuse the power of such knowledge, leading to discrimination based on a person's genetics, human genetic engineering, and other ethical concerns. This information could also lead to legal issues regarding health and privacy.

This page titled [4.3: Whole-Genome Sequencing](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

4.4: Genomics and Proteomics

Proteins are the final products of genes, which help perform the function that the gene encodes. Amino acids comprise proteins and play important roles in the cell. All enzymes (except ribozymes) are proteins that act as catalysts to 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 that a cell type produces. We can study proteomes using the knowledge of genomes because genes code for mRNAs, and the mRNAs encode proteins. Although mRNA analysis is a step in the right direction, not all mRNAs are translated into proteins. Proteomics is the study of proteomes' function. Proteomics complements genomics and is useful when scientists want to test their hypotheses that they based on genes. Even though all multicellular organisms' cells 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 alternately spliced (cut and pasted to create novel combinations and novel proteins) and many proteins modify themselves after translation by processes such as proteolytic cleavage, phosphorylation, glycosylation, and ubiquitination. There are also protein-protein interactions, which complicate studying proteomes. Although the genome provides a blueprint, the final architecture depends on several factors that can change the progression of events that generate the proteome.

Metabolomics is related to genomics and proteomics. **Metabolomics** involves studying small molecule metabolites in an organism. The metabolome is the complete set of metabolites that are related to an organism's genetic makeup. Metabolomics offers an opportunity to compare genetic makeup and physical characteristics, as well as genetic makeup and environmental factors. The goal of metabolome research is to identify, quantify, and catalogue all the metabolites in living organisms' tissues and fluids.

Basic Techniques in Protein Analysis

The ultimate goal of proteomics is to identify or compare the proteins expressed from a given genome under specific conditions, study the interactions between the proteins, and use the information to predict cell behavior or develop drug targets. Just as scientists analyze the genome using the basic DNA sequencing technique, proteomics requires techniques for protein analysis. The basic technique for protein analysis, analogous to DNA sequencing, is mass spectrometry. Mass spectrometry identifies and determines a molecule's characteristics. Advances in spectrometry have allowed researchers to analyze very small protein samples. X-ray crystallography, for example, enables scientists to determine a protein crystal's three-dimensional structure at atomic resolution. Another protein imaging technique, nuclear magnetic resonance (NMR), uses atoms' magnetic properties to determine the protein's three-dimensional structure in aqueous solution. Scientists have also used protein microarrays to study protein interactions. Large-scale adaptations of the basic two-hybrid screen (Figure 9.14) have provided the basis for protein microarrays. Scientists use computer software to analyze the vast amount of data for proteomic analysis.

Genomic- and proteomic-scale analyses are part of systems biology, which is the study of whole biological systems (genomes and proteomes) based on interactions within the system. The European Bioinformatics Institute and the Human Proteome Organization (HUPO) are developing and establishing effective tools to sort through the enormous pile of systems biology data. Because proteins are the direct products of genes and reflect activity at the genomic level, it is natural to use proteomes to compare the protein profiles of different cells to identify proteins and genes involved in disease processes. Most pharmaceutical drug trials target proteins. Researchers use information that they obtain from proteomics to identify novel drugs and to understand their mechanisms of action.

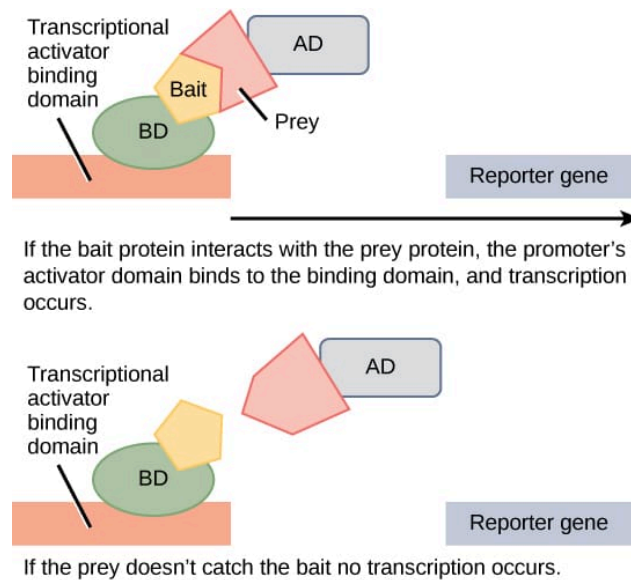


Figure 9.14 Scientists use two-hybrid screening to determine whether two proteins interact. In this method, a transcription factor splits into a DNA-binding domain (BD) and an activator domain (AD). The binding domain is able to bind the promoter in the activator domain's absence, but it does not turn on transcription. The bait protein attaches to the BD, and the prey protein attaches to the AD. Transcription occurs only if the prey “catches” the bait.

Scientists are challenged when implementing proteomic analysis because it is difficult to detect small protein quantities. Although mass spectrometry is good for detecting small protein amounts, variations in protein expression in diseased states can be difficult to discern. Proteins are naturally unstable molecules, which makes proteomic analysis much more difficult than genomic analysis.

Cancer Proteomics

Researchers are studying patients' genomes and proteomes to understand the genetic basis of diseases. The most prominent disease researchers are studying with proteomic approaches is cancer. These approaches improve screening and early cancer detection. Researchers are able to identify proteins whose expression indicates the disease process. An individual protein is a biomarker; whereas, a set of proteins with altered expression levels is a protein signature. For a biomarker or protein signature to be useful as a candidate for early cancer screening and detection, they must secrete in body fluids, such as sweat, blood, or urine, such that health professionals can perform large-scale screenings in a noninvasive fashion. The current problem with using biomarkers for early cancer detection is the high rate of false-negative results. A false negative is an incorrect test result that should have been positive. In other words, many cancer cases go undetected, which makes biomarkers unreliable. Some examples of protein biomarkers 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. Researchers are also using proteomics to develop individualized treatment plans, which involves predicting whether or not an individual will respond to specific drugs and the side effects that the individual may experience. Researchers also use proteomics to predict the possibility of disease recurrence.

The National Cancer Institute has developed programs to improve cancer detection and treatment. The Clinical Proteomic Technologies for Cancer and the Early Detection Research Network are efforts to identify protein signatures specific to different cancer types. The Biomedical Proteomics Program identifies protein signatures and designs effective therapies for cancer patients.

4.4.1: Biotechnology

Biotechnology is the use of biological agents for technological advancement. Biotechnology was used for breeding livestock and crops long before people understood the scientific basis of these techniques. Since the discovery of the structure of DNA in 1953, the biotechnology field has grown rapidly through both academic research and private companies. The primary applications of this technology are in medicine (vaccine and antibiotic production) and agriculture (crop genetic modification in order to increase yields). Biotechnology also has many industrial applications, such as fermentation, treating oil spills, and producing biofuels (Figure 9.1).

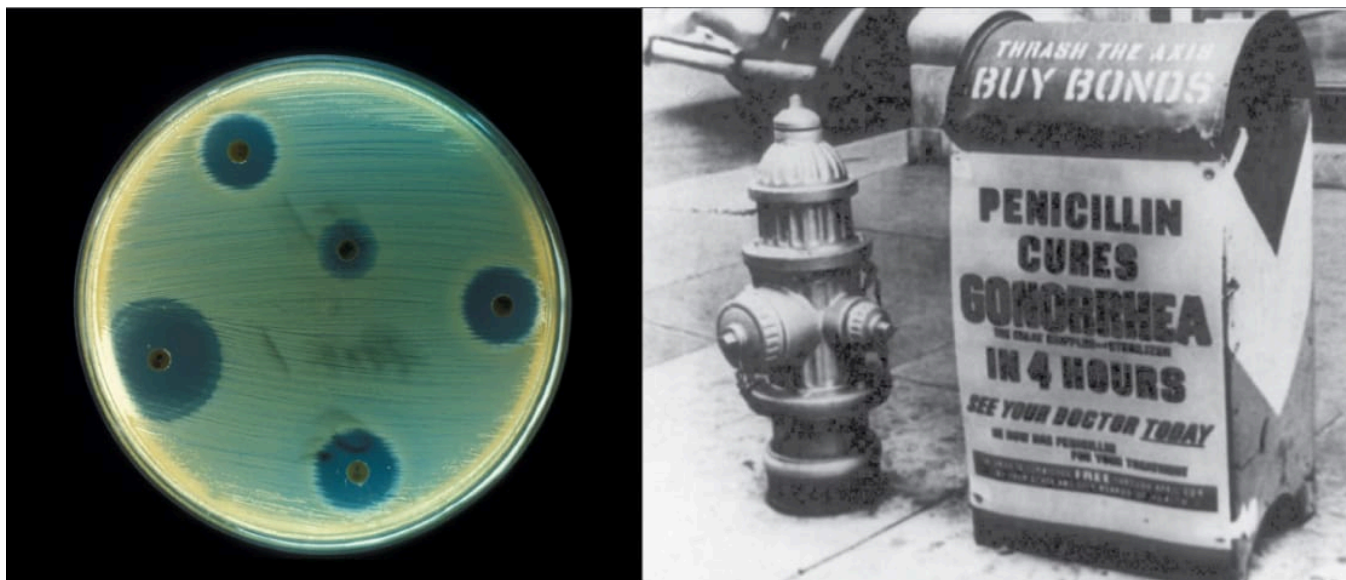


Figure 9.1 Fungi, bacteria, and other organisms that have antimicrobial properties produce antibiotics. The first antibiotic discovered was penicillin. Pharmaceutical companies now commercially produce and test antibiotics for their potential to inhibit bacterial growth. (credit "advertisement": modification of work by NIH; credit "test plate": modification of work by Don Stalons/CDC; scale-bar data from Matt Russell)

Basic Techniques to Manipulate Genetic Material (DNA and RNA)

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) linked by phosphodiester bonds. The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases. Exposure to high temperatures (DNA denaturation) can separate the two strands and cooling can reanneal them. The DNA polymerase enzyme can replicate the DNA. Unlike DNA, which is located in the eukaryotic cells' nucleus, RNA molecules leave the nucleus. The most common type of RNA that researchers analyze is the messenger RNA (mRNA) because it represents the protein-coding genes that are actively expressed. However, RNA molecules present some other challenges to analysis, as they are often less stable than DNA.

DNA and RNA Extraction

To study or manipulate nucleic acids, one must first isolate or extract the DNA or RNA from the cells. Researchers use various techniques to extract different types of DNA (Figure 9.2). Most nucleic acid extraction techniques involve steps to break open the cell and use enzymatic reactions to destroy all macromolecules that are not desired (such as unwanted molecule degradation and separation from the DNA sample). A lysis buffer (a solution which is mostly a detergent) breaks cells. Note that lysis means "to split". These enzymes break apart lipid molecules in the cell membranes and nuclear membranes. Enzymes such as proteases that break down proteins inactivate macromolecules, and ribonucleases (RNases) that break down RNA. Using alcohol precipitates the DNA. Human genomic DNA is usually visible as a gelatinous, white mass. One can store the DNA samples frozen at -80°C for

several years.

DNA Extraction

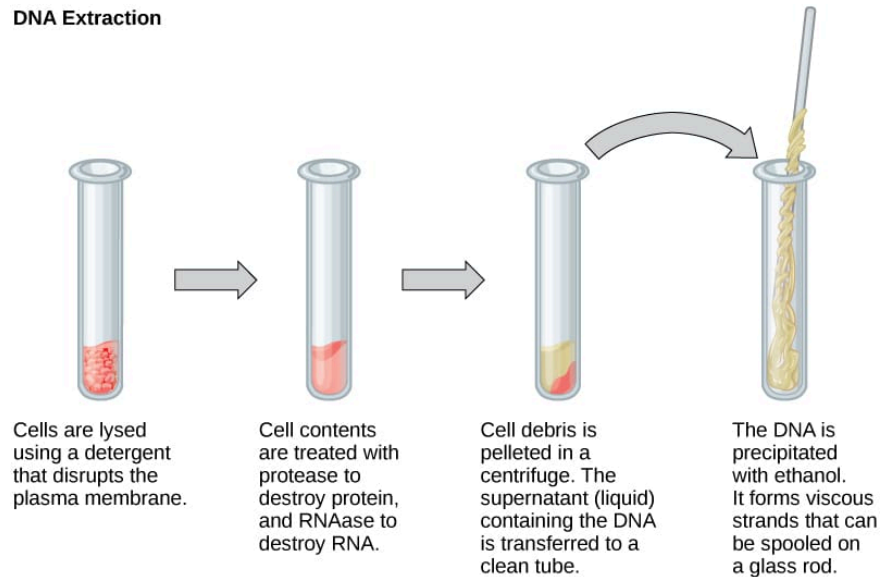


Figure 9.2 This diagram shows the basic method of DNA extraction.

Scientists perform RNA analysis to study gene expression patterns in cells. RNA is naturally very unstable because RNases are commonly present in nature and very difficult to inactivate. Similar to DNA, RNA extraction involves using various buffers and enzymes to inactivate macromolecules and preserve the RNA.

Gel Electrophoresis

Because nucleic acids are negatively charged ions at neutral or basic pH in an aqueous environment, an electric field can mobilize them. Gel electrophoresis is a technique that scientists use to separate molecules on the basis of size, using this charge. One can separate the nucleic acids as whole chromosomes or fragments. The nucleic acids load into a slot near the semisolid, porous gel matrix's negative electrode, and pulled toward the positive electrode at the gel's opposite end. Smaller molecules move through the gel's pores faster than larger molecules. This difference in the migration rate separates the fragments on the basis of size. There are molecular weight standard samples that researchers can run alongside the molecules to provide a size comparison. We can observe nucleic acids in a gel matrix using various fluorescent or colored dyes. Distinct nucleic acid fragments appear as bands at specific distances from the gel's top (the negative electrode end) on the basis of their size (Figure 9.3). A mixture of genomic DNA 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 gel's top.

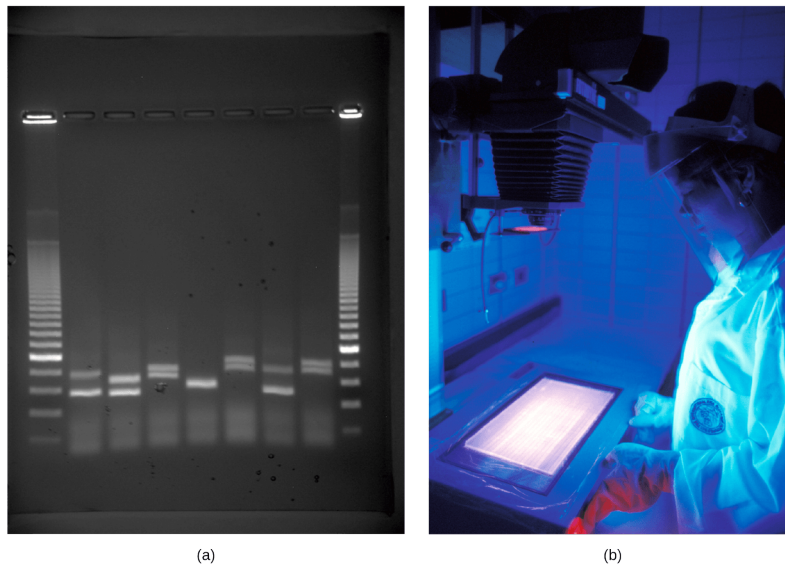


Figure 9.3 a) Shown are DNA fragments from seven samples run on a gel, stained with a fluorescent dye, and viewed under UV light; and b) a researcher from International Rice Research Institute, reviewing DNA profiles using UV light. (credit: a: James Jacob, Tompkins Cortland Community College b: International Rice Research Institute)

Nucleic Acid Fragment Amplification by Polymerase Chain Reaction

Although genomic DNA is visible to the naked eye when it is extracted in bulk, DNA analysis often requires focusing on one or more specific genome regions. **Polymerase chain reaction (PCR)** is a technique that scientists use to amplify specific DNA regions for further analysis (Figure 9.4). Researchers use PCR for many purposes in laboratories, such as cloning gene fragments to analyze genetic diseases, identifying contaminant foreign DNA in a sample, and amplifying DNA for sequencing. More practical applications include determining paternity and detecting genetic diseases.

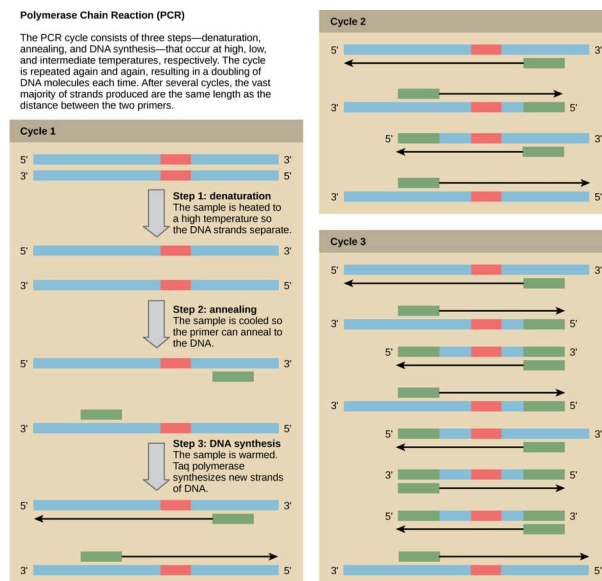


Figure 9.4 Scientists use polymerase chain reaction, or PCR, to amplify a specific DNA sequence. Primers—short pieces of DNA complementary to each end of the target sequence combine with genomic DNA, Taq polymerase, and deoxynucleotides. Taq polymerase is a DNA polymerase isolated from the thermostable bacterium *Thermus aquaticus* that is able to withstand the high temperatures that scientists use in PCR. *Thermus aquaticus* grows in the Lower Geyser Basin of Yellowstone National Park. Reverse transcriptase PCR (RT-PCR) is similar to PCR, but cDNA is made from an RNA template before PCR begins.

DNA fragments can also be amplified from an RNA template in a process called reverse transcriptase PCR (RT-PCR). The first step is to recreate the original DNA template strand (called cDNA) by applying DNA nucleotides to the mRNA. This process is

called **reverse transcription**. This requires the presence of an enzyme called reverse transcriptase. After the cDNA is made, regular PCR can be used to amplify it.

Link to Learning

Deepen your understanding of the polymerase chain reaction by watching [this video](#).

Hybridization, Southern Blotting, and Northern Blotting

Scientists can probe nucleic acid samples, such as fragmented genomic DNA and RNA extracts, for the presence of certain sequences. Scientists design and label short DNA fragments, or probes with radioactive or fluorescent dyes to aid detection. Gel electrophoresis separates the nucleic acid fragments according to their size. Scientists then transfer the fragments in the gel onto a nylon membrane in a procedure we call blotting (Figure 9.5). Scientists can then probe the nucleic acid fragments that are bound to the membrane's surface with specific radioactively or fluorescently labeled probe sequences. When scientists transfer DNA to a nylon membrane, they refer to the technique as Southern blotting. When they transfer the RNA to a nylon membrane, they call it Northern blotting. Scientists use Southern blots to detect the presence of certain DNA sequences in a given genome, and Northern blots to detect gene expression.

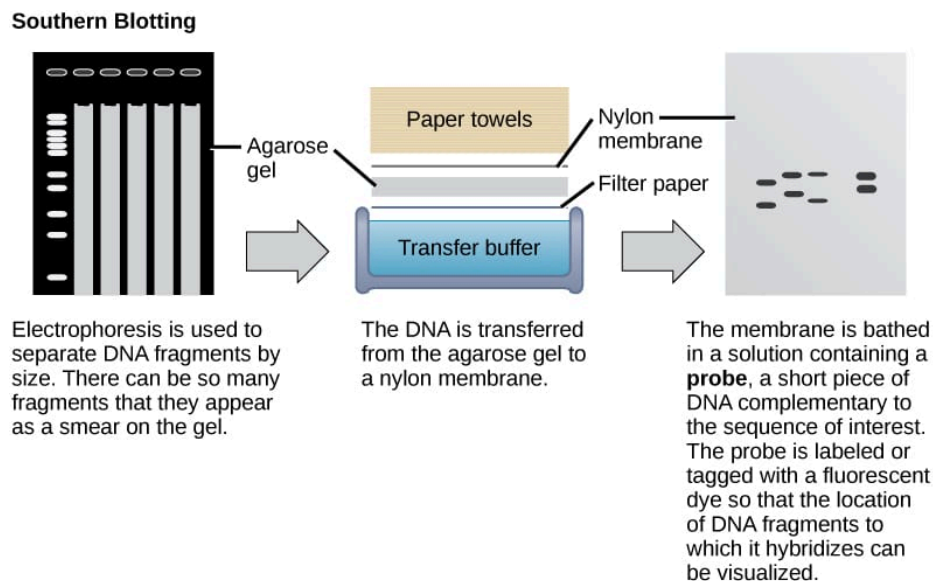


Figure 9.5 Scientists use Southern blotting to find a particular sequence in a DNA sample. Scientists separate DNA fragments on a gel, transfer them to a nylon membrane, and incubate them with a DNA probe complementary to the sequence of interest. Northern blotting is similar to Southern blotting, but scientists run RNA on the gel instead of DNA. In Western blotting, scientists run proteins on a gel and detect them using antibodies.

Molecular Cloning

In general, the word “cloning” means the creation of a perfect replica; however, 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 reproduce desired regions or fragments of the genome, a process that is referred to as molecular cloning. The technique offered methods to create new medicines and to overcome difficulties with existing ones. When Lydia Villa-Komaroff, working in the Gilbert Lab at Harvard, published the first paper outlining the technique for producing synthetic insulin, diabetes researchers and patients received new hope in fighting the disease. Insulin at that time was only produced using pig and cow pancreases, and the life-saving substance was often in short supply. Synthetic insulin, once mass produced, would solve that problem for many patients. These early discoveries led to the “BioTech Boom,” and spurred continued research and funding for newer and better ways to improve health.

Cloning small genome fragments allows researchers to manipulate and study specific genes (and their protein products), or

noncoding regions in isolation. A plasmid, or vector, is a small circular DNA molecule that replicates independently of the chromosomal DNA. In cloning, scientists can use the plasmid molecules to provide a "folder" in which to insert a desired DNA fragment. Plasmids are usually introduced into a bacterial host for proliferation. In the bacterial context, scientists call the DNA fragment from the human genome (or the genome of another studied organism) foreign DNA, or a transgene, to differentiate it from the bacterium's DNA, or 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). Scientists have repurposed and engineered plasmids as vectors for molecular cloning and the large-scale production of important reagents, such as insulin and human growth hormone. An important feature of plasmid vectors is the ease with which scientists can introduce a foreign DNA fragment via the multiple cloning site (MCS). The MCS is a short DNA sequence containing multiple sites that different commonly available restriction endonucleases can cut. **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 endonucleases make staggered cuts in the two DNA strands, such that the cut ends have a 2- or 4-base single-stranded overhang. Because these overhangs are capable of annealing with complementary overhangs, we call them **sticky ends**. Adding the enzyme DNA ligase permanently joins the DNA fragments via phosphodiester bonds. In this way, scientists can splice any DNA fragment generated by restriction endonuclease cleavage between the plasmid DNA's two ends that has been cut with the same restriction endonuclease (Figure 9.6).

Recombinant DNA Molecules

Plasmids with foreign DNA inserted into them are called **recombinant DNA molecules** because they are created artificially and do not occur in nature. They are also called chimeric molecules because the origin of different molecule parts of the molecules can be traced back to different species of biological organisms or even to chemical synthesis. We call proteins that are expressed from recombinant DNA molecules recombinant proteins. Not all recombinant plasmids are capable of expressing genes. The recombinant DNA may need to move into a different vector (or host) that is better designed for gene expression. Scientists may also engineer plasmids to express proteins only when certain environmental factors stimulate them, so they can control the recombinant proteins' expression.

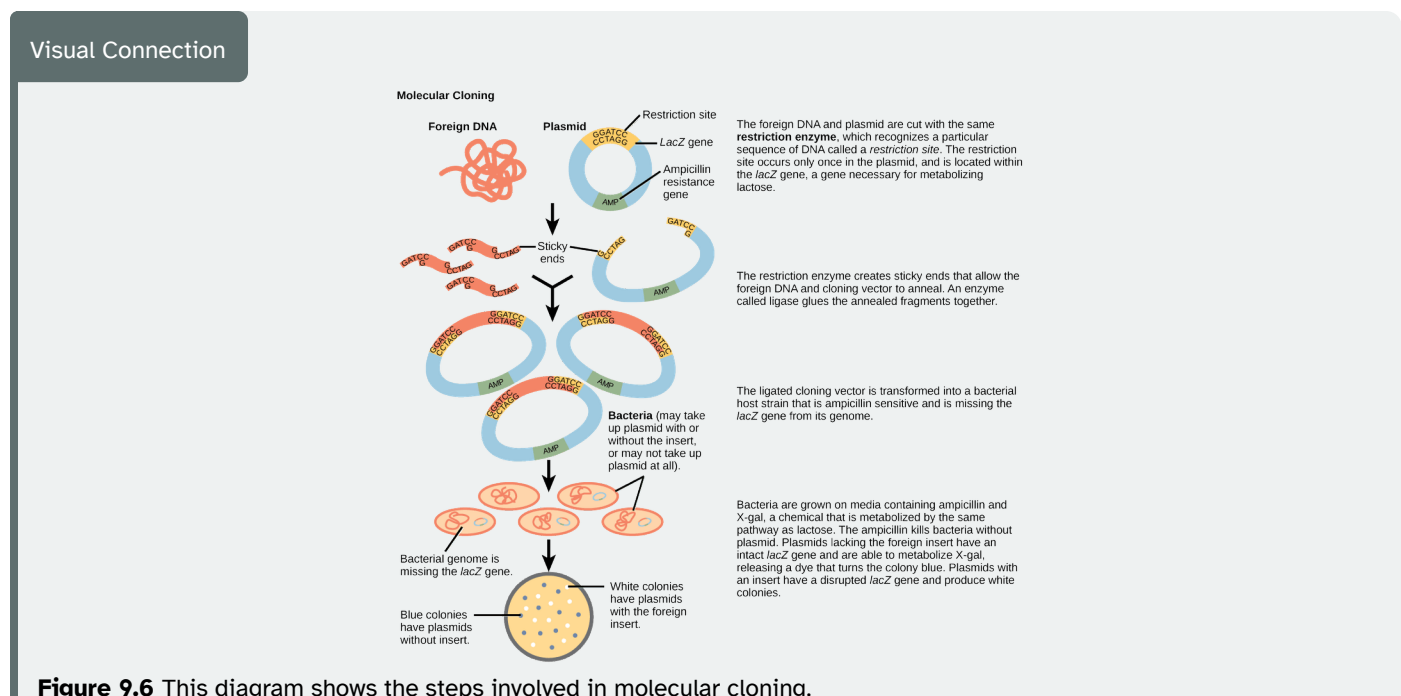


Figure 9.6 This diagram shows the steps involved in molecular cloning.

You are working in a molecular biology lab and, unbeknownst to you, your lab partner left the foreign genomic DNA that you are planning to clone on the lab bench overnight instead of storing it in the freezer. As a result, it was degraded by nucleases, but still used in the experiment. The plasmid, on the other hand, is fine. What results would you expect from your molecular cloning experiment?

- a. There will be no colonies on the bacterial plate.
- b. There will be blue colonies only.
- c. There will be blue and white colonies.
- d. There will be white colonies only.

Link to Learning

View an [animation of recombination in cloning](#) from the DNA Learning Center.

Cellular Cloning

Unicellular organisms, such as bacteria and yeast, naturally produce clones of themselves when they replicate asexually by binary fission; this is known as cellular cloning. The nuclear DNA duplicates by the process of mitosis, which creates an exact replica of the genetic material.

Reproductive Cloning

Reproductive cloning is a method scientists use to clone or identically copy an entire multicellular organism. Most multicellular organisms undergo reproduction by sexual means, which involves genetic hybridization of 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 artificially induce mammal asexual reproduction in the laboratory.

Parthenogenesis, or “virgin birth,” occurs when an embryo grows and develops without egg fertilization. This is a form of asexual reproduction. An example of parthenogenesis occurs in species in which the female lays an egg and if the egg is fertilized, it is a diploid egg and the individual develops into a female. If the egg is not fertilized, it remains a haploid egg and develops into a male. The unfertilized egg is a parthenogenic, or virgin egg. Some insects and reptiles lay parthenogenic eggs that can develop into adults.

Sexual reproduction requires two cells. When the haploid egg and sperm cells fuse, a diploid zygote results. The zygote nucleus contains the genetic information to produce a new individual. However, early embryonic development requires the cytoplasmic material contained in the egg cell. This idea forms the basis for reproductive cloning. Therefore, if we replace the egg cell's haploid nucleus with a diploid nucleus from the cell of any individual of the same species (a donor), it will become a zygote that is genetically identical to the donor. Somatic cell nuclear transfer is the technique of transferring a diploid nucleus into an enucleated egg. Scientists can use it for either therapeutic cloning or reproductive cloning.

The first cloned animal was Dolly, a sheep born in 1996. The reproductive cloning success rate at the time was very low. Dolly lived for seven years and died of respiratory complications (Figure 17.8). There is speculation that because the cell DNA belongs to an older individual, DNA's age may affect a cloned individual's life expectancy. Since Dolly, scientists have cloned successfully several animals such as horses, bulls, and goats, although these animals often exhibit facial, limb, and cardiac abnormalities. There have been attempts at producing cloned human embryos as sources of embryonic stem cells for therapeutic purposes. Therapeutic cloning produces stem cells in the attempt to remedy detrimental diseases or defects (unlike reproductive cloning, which aims to reproduce an organism). While some have met therapeutic cloning efforts with resistance because of bioethical considerations, new discoveries have identified additional opportunities for stem cell therapies. For example, Freda Miller and Elaine Fuchs, working independently, discovered stem cells in different layers of the skin. These cells help the skin repair itself, and their discovery may have applications in treatments of skin disease and potentially other conditions, such as nerve damage.

Visual Connection

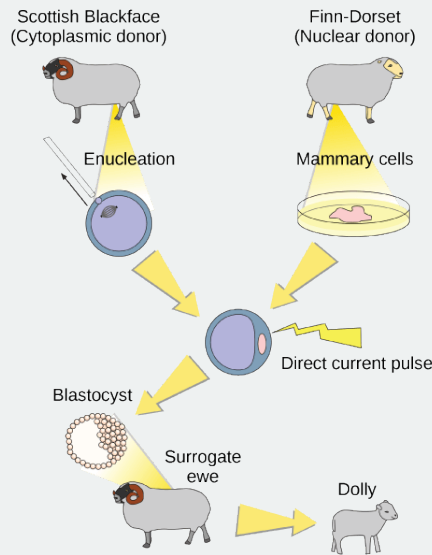


Figure 9.7 Dolly the sheep was the first mammal to be cloned. To create Dolly, they removed the nucleus from a donor egg cell. They then introduced the nucleus from a second sheep into the cell, which divided to the blastocyst stage before they implanted it in a surrogate mother. (credit: modification of work by "Squidoni"/Wikimedia Commons)

Do you think Dolly was a Finn-Dorset or a Scottish Blackface sheep?

Genetic Engineering

Genetic engineering is the alteration of an organism's genotype using recombinant DNA technology to modify an organism's DNA to achieve desirable traits. The addition of foreign DNA in the form of recombinant DNA vectors generated by molecular cloning is the most common method of genetic engineering. The organism that receives the recombinant DNA is a **genetically modified organism (GMO)**. If the foreign DNA comes from a different species, the host organism is **transgenic**. Scientists have genetically modified bacteria, plants, and animals since the early 1970s for academic, medical, agricultural, and industrial purposes. In the US, GMOs such as Roundup-ready soybeans and borer-resistant corn are part of many common processed foods.

Gene Targeting

Although classical methods of studying gene function 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, reverse genetics, has resulted in reversing the classic genetic methodology. This method would be similar 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 classical genetic method would compare insects that cannot fly with insects that can fly, and observe that the non-flying insects have lost wings. Similarly, mutating or deleting genes provides researchers with clues about gene function. We collectively call the methods they use to disable gene function gene targeting. Gene targeting is the use of recombinant DNA vectors to alter a particular gene's expression, either by introducing mutations in a gene, or by eliminating a certain gene's expression by deleting a part or all of the gene sequence from the organism's genome.

Biotechnology in Medicine and 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 the disease. Biotechnology in agriculture can enhance resistance to disease, pest, and environmental stress, and improve both crop yield and quality.

Genetic Diagnosis and Gene Therapy

Scientists call the process of testing for suspected genetic defects before administering treatment genetic diagnosis by genetic testing. Depending on the inheritance patterns of a disease-causing gene, family members are advised to undergo genetic testing. For example, doctors usually advise people diagnosed with breast cancer to have a biopsy so that the medical team can determine the genetic basis of cancer development. Doctors base treatment plans on genetic test findings that determine the type of cancer. If inherited gene mutations cause the cancer, doctors also advise other female relatives to undergo genetic testing and periodic screening for breast cancer. Doctors also offer genetic testing for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

Gene therapy is a genetic engineering technique used to cure disease. In its simplest form, it involves the introduction of a good gene at a random location in the genome to aid the cure of a disease that is caused by a mutated gene. The good gene is usually introduced into diseased cells as part of a vector transmitted by a virus that can infect the host cell and deliver the foreign DNA (Figure 9.8). More advanced forms of gene therapy try to correct the mutation at the original site in the genome, such as is the case with treatment of severe combined immunodeficiency (SCID).

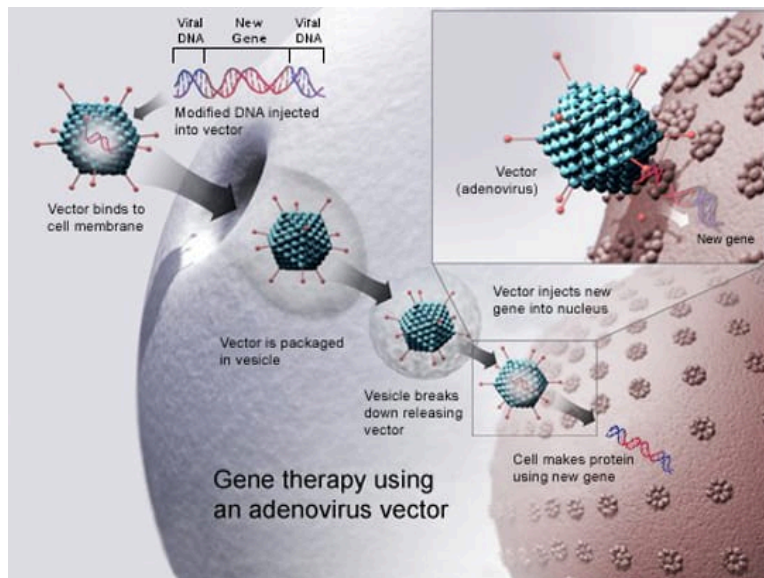


Figure 9.8 Gene therapy using an adenovirus vector can be used to cure certain genetic diseases in which a person has a defective gene. (credit: NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms to mount the initial immune response. Modern techniques use the genes of microorganisms cloned into vectors to mass produce the desired antigen. Doctors then introduce the antigen into the body to stimulate the primary immune response and trigger immune memory. The medical field has used genes cloned from the influenza virus to combat the constantly changing strains of this virus.

Antibiotics are a biotechnological product. Microorganisms, such as fungi, naturally produce them to attain an advantage over bacterial populations. Cultivating and manipulating fungal cells produces antibiotics.

Scientists used recombinant DNA technology to produce large-scale quantities of human insulin in *E. coli* as early as 1978. Previously, it was only possible to treat diabetes with pig insulin, which caused allergic reactions in humans because of differences in the gene product. In addition, doctors use human growth hormone (HGH) to treat growth disorders in children. Researchers cloned the HGH gene from a cDNA library and inserted it into *E. coli* cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins in medicine are successfully produced in bacteria, some proteins require a eukaryotic animal host for proper processing. For this reason, the desired genes are cloned and expressed in animals, such as sheep, goats,

chickens, and mice. We call animals that have been modified to express recombinant DNA transgenic animals. Several human proteins are expressed in transgenic sheep and goat milk, and some are expressed in chicken eggs. Scientists have used mice extensively for expressing and studying recombinant gene and mutation effects.

Transgenic Plants

Manipulating the DNA of plants (i.e., creating GMOs) has helped to create desirable traits, such as disease resistance, herbicide and pesticide resistance, better nutritional value, and better shelf-life (Figure 9.9). 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 9.9 Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

We call plants that have received recombinant DNA from other species transgenic plants. Because they are not natural, government agencies closely monitor transgenic plants and other GMOs 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, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants that scientists genetically engineered.

Transformation of Plants Using *Agrobacterium tumefaciens*

Gene transfer occurs naturally between species in microbial populations. Many viruses that cause human diseases, such as cancer, act by incorporating their DNA into the human genome. In plants, tumors caused by the bacterium *Agrobacterium tumefaciens* occur by DNA transfer from the bacterium to the plant. Although the tumors do not kill the plants, they stunt the plants and they become more susceptible to harsh environmental conditions. *A. tumefaciens* affects many plants such as walnuts, grapes, nut trees, and beets. Artificially introducing 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, Ti plasmids (tumor-inducing plasmids), that contain genes to produce tumors in plants. DNA from the Ti plasmid integrates into the infected plant cell's genome. Researchers manipulate the Ti plasmids to remove the tumor-causing genes and insert the desired DNA fragment for transfer into the plant genome. The Ti plasmids carry antibiotic resistance genes to aid selection and researchers can propagate them in *E. coli* cells as well.

The Organic Insecticide *Bacillus thuringiensis*

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals during sporulation that are toxic to many insect species that affect plants. Insects need to ingest Bt toxin in order to activate the toxin. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin activates in the insects' intestines, they die within a couple of days. Modern biotechnology has allowed plants to encode their own crystal Bt toxin that acts against insects. Scientists have cloned the crystal toxin genes from Bt and introduced them into plants. Bt toxin is safe for the environment, nontoxic to humans and other mammals, and organic farmers have approved it as a natural insecticide.

Flavr Savr Tomato

The first GM crop on the market was the Flavr Savr Tomato in 1994. Scientists used antisense RNA technology to slow the softening and rotting process caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the tomato's flavor. The Flavr Savr tomato did not successfully stay in the market because of problems maintaining and shipping the crop.

Key Terms

Biotechnology

The use of living organisms, cells, or biological molecules to develop technologies and products that improve medicine, agriculture, industry, and research.

DNA Extraction

The process of isolating DNA from cells for analysis or manipulation.

RNA Extraction

The process of isolating RNA from cells while preventing RNA degradation.

Lysis Buffer

A solution containing detergents and other chemicals used to break open cells during DNA or RNA extraction.

Gel Electrophoresis

A laboratory technique that separates DNA, RNA, or proteins according to size by applying an electric field to a gel matrix.

DNA Denaturation

The separation of the two strands of DNA by breaking the hydrogen bonds between complementary bases, typically through heating.

Polymerase Chain Reaction (PCR)

A laboratory technique used to amplify a specific DNA sequence, producing millions of copies.

Primer

A short DNA sequence that binds to a target DNA region and provides a starting point for DNA polymerase during PCR.

Taq Polymerase

A heat-stable DNA polymerase isolated from *Thermus aquaticus* that is commonly used during PCR.

Reverse Transcriptase PCR (RT-PCR)

A technique in which RNA is first converted into complementary DNA (cDNA) before PCR amplification.

Complementary DNA (cDNA)

A DNA molecule synthesized from an RNA template using the enzyme reverse transcriptase.

DNA Probe

A short, labeled DNA fragment used to detect complementary DNA or RNA sequences.

Southern Blot

A technique used to detect specific DNA sequences by transferring DNA fragments from a gel to a membrane and probing them

with labeled DNA.

Northern Blot

A technique used to detect specific RNA molecules and analyze gene expression.

Plasmid

A small, circular DNA molecule that replicates independently of the bacterial chromosome and is commonly used as a cloning vector.

Vector

A DNA molecule used to carry foreign DNA into a host cell.

Restriction Endonuclease (Restriction Enzyme)

An enzyme that recognizes specific DNA sequences and cuts DNA at or near those sites.

Sticky Ends

Single-stranded DNA overhangs produced by certain restriction enzymes that facilitate the joining of DNA fragments.

Molecular Cloning

The process of inserting a DNA fragment into a vector so it can be replicated and studied.

Recombinant DNA

DNA artificially created by combining DNA from two or more different sources.

Genetically Modified Organism (GMO)

An organism whose genetic material has been altered using biotechnology.

Transgenic Organism

A genetically modified organism that contains DNA from a different species.

Genetic Engineering

The deliberate modification of an organism's DNA using recombinant DNA technology.

Gene Targeting

A technique used to alter or disrupt a specific gene in order to study its function.

Genetic Testing

The analysis of DNA to identify inherited genetic variants associated with disease.

Gene Therapy

A medical technique that treats disease by introducing or modifying genes within a patient's cells.

Transgenic Plant

A plant that has received DNA from another organism through genetic engineering.

Check Your Understanding

1. What is biotechnology, and how has it benefited medicine and agriculture?
2. Why has biotechnology advanced rapidly since the discovery of DNA's structure?
3. Why must DNA or RNA first be extracted before molecular analysis can be performed?
4. What is the function of a lysis buffer during nucleic acid extraction?
5. Why is RNA generally more difficult to work with than DNA?
6. Explain how gel electrophoresis separates DNA fragments.
7. What is the purpose of PCR?
8. Why is Taq polymerase essential for PCR?
9. What role do primers play during PCR?
10. How does RT-PCR differ from standard PCR?
11. Compare Southern blotting and Northern blotting.
12. What is the function of a DNA probe?

13. What is a plasmid, and why is it useful in molecular cloning?
14. What is the function of restriction enzymes?
15. What are sticky ends, and why are they important during cloning?
16. Why are bacteria commonly used in molecular cloning experiments?
17. What is the difference between a genetically modified organism (GMO) and a transgenic organism?
18. How is recombinant DNA technology used to produce human insulin?

INDEX

A

- ABO blood group
 - 3.1.7: White spotting - when there's more than two alleles
- activator
 - 2.3.2: Prokaryotic Gene Regulation
- allele
 - 3.1.2: Laws of Inheritance
 - 3.1.4: Black fur color - a dominant trait
- aminoacyl tRNA synthetase
 - 2.2.2: Ribosomes and Protein Synthesis
- anaphase
 - 2.1.1: The Cell Cycle
- aneuploid
 - 2.1.6: Errors in Meiosis
- anneal
 - 4.1: Biotechnology
- antibiotic resistance
 - 4.4.1: Biotechnology
- anticodon
 - 1.3.4: RNA Processing
- autosome
 - 2.1.6: Errors in Meiosis
 - 3.1.9: Hemophilia - a sex-linked disorder

B

- binary fission
 - 2.1.4: Prokaryotic Cell Division
- biomarker
 - 4.4: Genomics and Proteomics
- biotechnology
 - 4.1: Biotechnology
 - 4.4.1: Biotechnology
- brindle
 - 3.1.6: Brindle color - partial dominance and epistasis

C

- CAAT box
 - 1.3.3: Eukaryotic Transcription
- cancer
 - 2.3.4: Cancer and Gene Regulation
- CAP
 - 2.3.2: Prokaryotic Gene Regulation
- catabolite activator protein
 - 2.3.2: Prokaryotic Gene Regulation
- cDNA library
 - 4.2: Mapping Genomes
- cell cycle
 - 1.1.3: Genome Organization
 - 2.1.1: The Cell Cycle
- cell cycle checkpoint
 - 2.1.2: Control of the Cell Cycle
- cell plate
 - 2.1.1: The Cell Cycle
- cellular cloning
 - 4.4.1: Biotechnology
- central dogma
 - 1.3.1: Transcription
 - 2.2.1: The Genetic Code

- centriole
 - 2.1.1: The Cell Cycle
- centromere
 - 1.1.3: Genome Organization
- chain termination method
 - 4.3: Whole-Genome Sequencing
- Chargaff's rules
 - 1.1.2: The Structure of DNA
- chiasmata
 - 2.1.5: Meiosis
- chromatid
 - 1.1.3: Genome Organization
- chromosome inversion
 - 2.1.6: Errors in Meiosis
- cleavage furrow
 - 2.1.1: The Cell Cycle
- clone
 - 4.4.1: Biotechnology
- cloning
 - 4.1: Biotechnology
- codominance
 - 3.1.3: Extensions of the Laws of Inheritance
- codon
 - 2.2.1: The Genetic Code
- colinear
 - 2.2.1: The Genetic Code
- color blindness
 - 3.1.9: Hemophilia - a sex-linked disorder
- condensin
 - 2.1.1: The Cell Cycle
- consensus
 - 1.3.2: Bacterial Transcription
- contig
 - 4.3: Whole-Genome Sequencing
- continuous variation
 - 3.1.1: Mendel's Experiments
- core enzyme
 - 1.3.2: Bacterial Transcription
- crossing over
 - 2.1.5: Meiosis
- cyclin
 - 2.1.2: Control of the Cell Cycle
- cytogenetic mapping
 - 4.2: Mapping Genomes
- cytokinesis
 - 2.1.1: The Cell Cycle

D

- degeneracy
 - 2.2.1: The Genetic Code
- deletion mutation
 - 2.4.2: Types of Mutations
- deoxynucleotide
 - 4.3: Whole-Genome Sequencing
- deoxyribose
 - 1.1.2: The Structure of DNA
- dideoxynucleotide
 - 4.3: Whole-Genome Sequencing

- dihybrid
 - 3.1.2: Laws of Inheritance
- diploid
 - 1.1.3: Genome Organization
- discontinuous variation
 - 3.1.1: Mendel's Experiments
- dispersive model
 - 1.2.1: Basics of DNA Replication
- DNA
 - 1.1.1: Historical Basis of Modern Understanding
- DNA ligase
 - 1.2.2: DNA Replication
 - 1.2.3: DNA Replication in Bacteria
- DNA methylation
 - 2.3.4: Cancer and Gene Regulation
- DNA microarray
 - 4.3: Whole-Genome Sequencing
- DNA polymerase
 - 1.2.2: DNA Replication
- DNA polymerase III
 - 1.2.3: DNA Replication in Bacteria
- DNA replication
 - 1.2.1: Basics of DNA Replication
 - 1.2.4: DNA Replication in Eukaryotes
- dominant
 - 3.1.1: Mendel's Experiments
 - 3.1.4: Black fur color - a dominant trait
- double helix
 - 1.1.2: The Structure of DNA
- downstream
 - 1.3.2: Bacterial Transcription

E

- embryonic lethality
 - 3.1.8: KIT - embryonic lethality
- environmental factor
 - 3.1.10: It's not all in the genes - the effect of environment
- epigenetic
 - 2.3.1: Regulation of Gene Expression
- Epigenetic mechanisms
 - 2.3.3: Eukaryotic Gene Regulation
- epistasis
 - 3.1.3: Extensions of the Laws of Inheritance
 - 3.1.6: Brindle color - partial dominance and epistasis
- EST
 - 4.2: Mapping Genomes
- eukaryotes
 - 1.2.4: DNA Replication in Eukaryotes
- euploid
 - 2.1.6: Errors in Meiosis
- exon
 - 1.3.1: Transcription
 - 1.3.4: RNA Processing
- expressed sequence tag
 - 4.2: Mapping Genomes

F

- F1**
 - [3.1.1: Mendel's Experiments](#)
- F2**
 - [3.1.1: Mendel's Experiments](#)
- facilitates chromatin transcription**
 - [1.3.3: Eukaryotic Transcription](#)
- FACT**
 - [1.3.3: Eukaryotic Transcription](#)
- false negative**
 - [4.4: Genomics and Proteomics](#)
- fertilization**
 - [2.1.5: Meiosis](#)
- foreign DNA**
 - [4.4.1: Biotechnology](#)
- frameshift mutation**
 - [2.4.2: Types of Mutations](#)
- Frederick Griffith**
 - [1.1.1: Historical Basis of Modern Understanding](#)
- Friedrich Miescher**
 - [1.1.1: Historical Basis of Modern Understanding](#)
- FtsZ**
 - [2.1.4: Prokaryotic Cell Division](#)

G

- G0 phase**
 - [2.1.1: The Cell Cycle](#)
- G1 phase**
 - [2.1.1: The Cell Cycle](#)
- G2 phase**
 - [2.1.1: The Cell Cycle](#)
- gamete**
 - [1.1.3: Genome Organization](#)
- gel electrophoresis**
 - [4.1: Biotechnology](#)
 - [4.4.1: Biotechnology](#)
- gene**
 - [1.1.3: Genome Organization](#)
- gene duplication**
 - [2.4.2: Types of Mutations](#)
- gene expression**
 - [2.3.1: Regulation of Gene Expression](#)
- gene regulation**
 - [2.3.3: Eukaryotic Gene Regulation](#)
 - [2.3.4: Cancer and Gene Regulation](#)
- gene targeting**
 - [4.4.1: Biotechnology](#)
- gene therapy**
 - [4.4.1: Biotechnology](#)
- genetic diagnosis**
 - [4.4.1: Biotechnology](#)
- genetic engineering**
 - [4.1: Biotechnology](#)
 - [4.4.1: Biotechnology](#)
- genetic map**
 - [4.2: Mapping Genomes](#)
- genetic marker**
 - [4.2: Mapping Genomes](#)
- genetic recombination**
 - [4.2: Mapping Genomes](#)
- genetic testing**
 - [4.4.1: Biotechnology](#)
- genetically modified organism**
 - [4.4.1: Biotechnology](#)

genetically modified organism (GMO)

[4.1: Biotechnology](#)

genome

[1.1.3: Genome Organization](#)

genome annotation

[4.3: Whole-Genome Sequencing](#)

genome mapping

[4.2: Mapping Genomes](#)

genomic library

[4.2: Mapping Genomes](#)

genomics

[4.2: Mapping Genomes](#)

genotype

[3.1.2: Laws of Inheritance](#)

GMO

[4.4.1: Biotechnology](#)

H

- hairpin**
 - [1.3.2: Bacterial Transcription](#)
- haploid**
 - [1.1.3: Genome Organization](#)
- helicase**
 - [1.2.2: DNA Replication](#)
 - [1.2.3: DNA Replication in Bacteria](#)
- hemizygous**
 - [3.1.3: Extensions of the Laws of Inheritance](#)
- hemophilia**
 - [3.1.9: Hemophilia - a sex-linked disorder](#)
- heterozygous**
 - [3.1.2: Laws of Inheritance](#)
- histone**
 - [1.1.3: Genome Organization](#)
- histone acetylation**
 - [2.3.4: Cancer and Gene Regulation](#)
- holoenzyme**
 - [1.3.2: Bacterial Transcription](#)
- homologous chromosomes**
 - [1.1.3: Genome Organization](#)
- homozygous**
 - [3.1.2: Laws of Inheritance](#)
- host DNA**
 - [4.4.1: Biotechnology](#)
- hybridization**
 - [3.1.1: Mendel's Experiments](#)

I

- incomplete dominance**
 - [3.1.3: Extensions of the Laws of Inheritance](#)
- inducible operon**
 - [2.3.2: Prokaryotic Gene Regulation](#)
- initiation site**
 - [1.3.2: Bacterial Transcription](#)
- initiator tRNA**
 - [2.2.2: Ribosomes and Protein Synthesis](#)
- insertion mutation**
 - [2.4.2: Types of Mutations](#)
- interkinesis**
 - [2.1.5: Meiosis](#)
- interphase**
 - [2.1.1: The Cell Cycle](#)
- intron**
 - [1.3.1: Transcription](#)
 - [1.3.4: RNA Processing](#)

K

- karyogram**
 - [2.1.6: Errors in Meiosis](#)
- karyotype**
 - [2.1.6: Errors in Meiosis](#)
- kinetochore**
 - [2.1.1: The Cell Cycle](#)
- Kozak's rules**
 - [2.2.2: Ribosomes and Protein Synthesis](#)

L

- lac operon**
 - [2.3.2: Prokaryotic Gene Regulation](#)
- lagging strand**
 - [1.2.2: DNA Replication](#)
 - [1.2.3: DNA Replication in Bacteria](#)
- law of dominance**
 - [3.1.2: Laws of Inheritance](#)
- Law of Independent Assortment**
 - [3.1.2: Laws of Inheritance](#)
- Law of Segregation**
 - [3.1.2: Laws of Inheritance](#)
- leading strand**
 - [1.2.2: DNA Replication](#)
 - [1.2.3: DNA Replication in Bacteria](#)
- linkage**
 - [3.1.3: Extensions of the Laws of Inheritance](#)
- linkage analysis**
 - [4.2: Mapping Genomes](#)
- locus**
 - [1.1.3: Genome Organization](#)
- lysis buffer**
 - [4.4.1: Biotechnology](#)

M

- MC1R**
 - [3.1.11: Pleiotropy - one gene affects more than one trait](#)
- MCS**
 - [4.4.1: Biotechnology](#)
- meiosis I**
 - [2.1.5: Meiosis](#)
- meiosis II**
 - [2.1.5: Meiosis](#)
- metabolome**
 - [4.4: Genomics and Proteomics](#)
- metabolomics**
 - [4.4: Genomics and Proteomics](#)
- metaphase**
 - [2.1.1: The Cell Cycle](#)
- metaphase plate**
 - [2.1.1: The Cell Cycle](#)
- microsatellite polymorphism**
 - [4.2: Mapping Genomes](#)
- mismatch repair**
 - [1.2.2: DNA Replication](#)
- missense mutation**
 - [2.4.2: Types of Mutations](#)
- mitosis**
 - [2.1.1: The Cell Cycle](#)
- mitotic phase**
 - [2.1.1: The Cell Cycle](#)
- mitotic spindle**
 - [2.1.1: The Cell Cycle](#)

model organism

[4.3: Whole-Genome Sequencing](#)

model system

[3.1.1: Mendel's Experiments](#)

molecular cloning

[4.4.1: Biotechnology](#)

monohybrid

[3.1.2: Laws of Inheritance](#)

monosomy

[2.1.6: Errors in Meiosis](#)

mRNA

[1.3.1: Transcription](#)

multiple allele traits

[3.1.7: White spotting - when there's more than two alleles](#)

multiple cloning site

[4.4.1: Biotechnology](#)

mutation

[1.2.2: DNA Replication](#)

myc

[2.3.4: Cancer and Gene Regulation](#)

N

negative regulator

[2.3.2: Prokaryotic Gene Regulation](#)

nitrogenous base

[1.1.2: The Structure of DNA](#)

nondisjunction

[2.1.6: Errors in Meiosis](#)

nonsense codon

[2.2.1: The Genetic Code](#)

nonsense mutation

[2.4.2: Types of Mutations](#)

nontemplate strand

[1.3.1: Transcription](#)
[1.3.2: Bacterial Transcription](#)

northern blotting

[4.4.1: Biotechnology](#)

nucleosome

[1.1.3: Genome Organization](#)

nucleotide excision repair

[1.2.2: DNA Replication](#)

O

Octamer box

[1.3.3: Eukaryotic Transcription](#)

Okazaki fragment

[1.2.3: DNA Replication in Bacteria](#)

Okazaki fragments

[1.2.2: DNA Replication](#)

oncogene

[2.1.3: Cancer and the Cell Cycle](#)

operator

[2.3.2: Prokaryotic Gene Regulation](#)

operons

[2.3.2: Prokaryotic Gene Regulation](#)

ORI

[2.1.4: Prokaryotic Cell Division](#)

origin

[2.1.4: Prokaryotic Cell Division](#)

origin of replication

[1.2.3: DNA Replication in Bacteria](#)

P

P

[3.1.1: Mendel's Experiments](#)

p21

[2.1.2: Control of the Cell Cycle](#)

p53

[2.1.2: Control of the Cell Cycle](#)

partial dominance

[3.1.6: Brindle color - partial dominance and epistasis](#)

PCR

[4.4.1: Biotechnology](#)

peptidyl transferase

[2.2.2: Ribosomes and Protein Synthesis](#)

phenotype

[3.1.2: Laws of Inheritance](#)

phenylketonuria (PKU)

[3.1.11: Pleiotropy - one gene affects more than one trait](#)

phosphate group

[1.1.2: The Structure of DNA](#)

physical map

[4.2: Mapping Genomes](#)

plasmid

[1.3.2: Bacterial Transcription](#)
[4.1: Biotechnology](#)

pleiotropy

[3.1.11: Pleiotropy - one gene affects more than one trait](#)

polymerase chain reaction (PCR)

[4.1: Biotechnology](#)
[4.4.1: Biotechnology](#)

polyploid

[2.1.6: Errors in Meiosis](#)

polysome

[2.2.2: Ribosomes and Protein Synthesis](#)

positive regulator

[2.3.2: Prokaryotic Gene Regulation](#)

preinitiation complex

[1.3.3: Eukaryotic Transcription](#)

primer

[1.2.2: DNA Replication](#)
[1.2.3: DNA Replication in Bacteria](#)

probe

[4.4.1: Biotechnology](#)

prometaphase

[2.1.1: The Cell Cycle](#)

promoter

[1.3.1: Transcription](#)
[1.3.2: Bacterial Transcription](#)

prophase

[2.1.1: The Cell Cycle](#)

protease

[4.4.1: Biotechnology](#)

protein signature

[4.4: Genomics and Proteomics](#)

proteome

[4.4: Genomics and Proteomics](#)

proteomics

[4.4: Genomics and Proteomics](#)

Punnett square

[3.1.2: Laws of Inheritance](#)

Q

quiescent

[2.1.1: The Cell Cycle](#)

R

radiation hybrid mapping

[4.2: Mapping Genomes](#)

reading frame

[2.2.1: The Genetic Code](#)

recessive

[3.1.1: Mendel's Experiments](#)
[3.1.4: Black fur color - a dominant trait](#)

reciprocal cross

[3.1.1: Mendel's Experiments](#)

recombinant

[2.1.5: Meiosis](#)

recombinant DNA

[4.1: Biotechnology](#)
[4.4.1: Biotechnology](#)

recombinant protein

[4.1: Biotechnology](#)
[4.4.1: Biotechnology](#)

recombination

[3.1.3: Extensions of the Laws of Inheritance](#)

reduction division

[2.1.5: Meiosis](#)

repeat expansion

[2.4.2: Types of Mutations](#)

replication fork

[1.2.2: DNA Replication](#)
[1.2.3: DNA Replication in Bacteria](#)

repressor

[2.3.2: Prokaryotic Gene Regulation](#)

reproductive cloning

[4.1: Biotechnology](#)
[4.4.1: Biotechnology](#)

restriction endonuclease

[4.4.1: Biotechnology](#)

restriction enzyme

[4.1: Biotechnology](#)

restriction fragment length polymorphism

[4.2: Mapping Genomes](#)

retinoblastoma protein

[2.1.2: Control of the Cell Cycle](#)

reverse genetics

[4.1: Biotechnology](#)
[4.4.1: Biotechnology](#)

reverse transcriptase PCR

[4.4.1: Biotechnology](#)

RFLP

[4.2: Mapping Genomes](#)

ribonuclease

[4.4.1: Biotechnology](#)

RNA editing

[1.3.4: RNA Processing](#)

RNA polymerase

[1.3.1: Transcription](#)

RNA primase

[1.2.3: DNA Replication in Bacteria](#)

S

S phase

[2.1.1: The Cell Cycle](#)

semiconservative replication

[1.2.2: DNA Replication](#)

septum

[2.1.4: Prokaryotic Cell Division](#)

sequence mapping

[4.2: Mapping Genomes](#)

sex chromosome

[3.1.9: Hemophilia - a sex-linked disorder](#)

shotgun sequencing

[4.3: Whole-Genome Sequencing](#)

signal sequence

[2.2.2: Ribosomes and Protein Synthesis](#)

silent mutation

[2.4.2: Types of Mutations](#)

single nucleotide polymorphism

[4.2: Mapping Genomes](#)

sliding clamp

[1.2.3: DNA Replication in Bacteria](#)

small nuclear RNA

[1.3.3: Eukaryotic Transcription](#)

SNP

[4.2: Mapping Genomes](#)

somatic cell

[2.1.5: Meiosis](#)

Southern blotting

[4.4.1: Biotechnology](#)

splicing

[1.3.1: Transcription](#)

[1.3.4: RNA Processing](#)

start codon

[2.2.2: Ribosomes and Protein Synthesis](#)

synapsis

[2.1.5: Meiosis](#)

systems biology

[4.4: Genomics and Proteomics](#)

T

TATA box

[1.3.2: Bacterial Transcription](#)

telomerase

[1.2.2: DNA Replication](#)

[1.2.4: DNA Replication in Eukaryotes](#)

telomere

[1.2.2: DNA Replication](#)

[1.2.4: DNA Replication in Eukaryotes](#)

telophase

[2.1.1: The Cell Cycle](#)

template strand

[1.3.1: Transcription](#)

[1.3.2: Bacterial Transcription](#)

test cross

[3.1.2: Laws of Inheritance](#)

tetrad

[2.1.5: Meiosis](#)

Ti plasmid

[4.4.1: Biotechnology](#)

topoisomerase

[1.2.3: DNA Replication in Bacteria](#)

trait

[3.1.1: Mendel's Experiments](#)

transcription bubble

[1.3.1: Transcription](#)

transcription factor

[2.3.3: Eukaryotic Gene Regulation](#)

transformation

[1.1.1: Historical Basis of Modern Understanding](#)

transgenic

[4.1: Biotechnology](#)

[4.4.1: Biotechnology](#)

translocation

[2.1.6: Errors in Meiosis](#)

trisomy

[2.1.6: Errors in Meiosis](#)

Trp operon

[2.3.2: Prokaryotic Gene Regulation](#)

tryptophan

[2.3.2: Prokaryotic Gene Regulation](#)

tubulin

[2.1.4: Prokaryotic Cell Division](#)

tumor suppressor gene

[2.1.3: Cancer and the Cell Cycle](#)

U

upstream

[1.3.2: Bacterial Transcription](#)

V

variable number of tandem repeats

[4.2: Mapping Genomes](#)

VNTRs

[4.2: Mapping Genomes](#)

W

wild type

[3.1.3: Extensions of the Laws of Inheritance](#)

X

X inactivation

[2.1.6: Errors in Meiosis](#)

GLOSSARY

S

Sample Word 1 | Sample Definition 1

DETAILED LICENSING

Overview

Title: BIOS 220: Introduction to Genetics

Webpages: 63

Applicable Restrictions: Noncommercial

All licenses found:

- [CC BY 4.0](#): 68.3% (43 pages)
- [Undeclared](#): 20.6% (13 pages)
- [CC BY-NC-SA 4.0](#): 6.3% (4 pages)
- [Public Domain](#): 4.8% (3 pages)

By Page

- BIOS 220: Introduction to Genetics - [CC BY-NC-SA 4.0](#)
 - Front Matter - *Undeclared*
 - TitlePage - [CC BY-NC-SA 4.0](#)
 - InfoPage - [Public Domain](#)
 - Table of Contents - [Public Domain](#)
 - Licensing - [Public Domain](#)
 - 1: Module 1 - [CC BY-NC-SA 4.0](#)
 - 1.1: What is DNA? - [CC BY-NC-SA 4.0](#)
 - 1.1.1: Historical Basis of Modern Understanding - [CC BY 4.0](#)
 - 1.1.2: The Structure of DNA - [CC BY 4.0](#)
 - 1.1.3: Genome Organization - [CC BY 4.0](#)
 - 1.2: DNA Replication - [CC BY 4.0](#)
 - 1.2.1: Basics of DNA Replication - [CC BY 4.0](#)
 - 1.2.2: DNA Replication - [CC BY 4.0](#)
 - 1.2.3: DNA Replication in Bacteria - [CC BY 4.0](#)
 - 1.2.4: DNA Replication in Eukaryotes - [CC BY 4.0](#)
 - 1.3: Transcription - *Undeclared*
 - 1.3.1: Transcription - [CC BY 4.0](#)
 - 1.3.2: Bacterial Transcription - [CC BY 4.0](#)
 - 1.3.3: Eukaryotic Transcription - [CC BY 4.0](#)
 - 1.3.4: RNA Processing - [CC BY 4.0](#)
 - 2: Module 2 - *Undeclared*
 - 2.1: Mitosis, Meiosis, and the Cell Cycle - [CC BY 4.0](#)
 - 2.1.1: The Cell Cycle - [CC BY 4.0](#)
 - 2.1.2: Control of the Cell Cycle - [CC BY 4.0](#)
 - 2.1.3: Cancer and the Cell Cycle - [CC BY 4.0](#)
 - 2.1.4: Prokaryotic Cell Division - [CC BY 4.0](#)
 - 2.1.5: Meiosis - [CC BY 4.0](#)
 - 2.1.6: Errors in Meiosis - [CC BY 4.0](#)
 - 2.2: Translation - *Undeclared*
 - 2.2.1: The Genetic Code - [CC BY 4.0](#)
 - 2.2.2: Ribosomes and Protein Synthesis - [CC BY 4.0](#)
 - 2.3: Gene Expression - *Undeclared*
 - 2.3.1: Regulation of Gene Expression - [CC BY 4.0](#)
 - 2.3.2: Prokaryotic Gene Regulation - [CC BY 4.0](#)
 - 2.3.3: Eukaryotic Gene Regulation - [CC BY 4.0](#)
 - 2.3.4: Cancer and Gene Regulation - [CC BY 4.0](#)
 - 2.4: Mutations and DNA Repair - *Undeclared*
 - 2.4.1: DNA Damage and Repair - [CC BY 4.0](#)
 - 2.4.2: Types of Mutations - [CC BY 4.0](#)
 - 3: Module 3 - *Undeclared*
 - 3.1: Inheritance - *Undeclared*
 - 3.1.1: Mendel's Experiments - [CC BY 4.0](#)
 - 3.1.2: Laws of Inheritance - [CC BY 4.0](#)
 - 3.1.3: Extensions of the Laws of Inheritance - [CC BY 4.0](#)
 - 3.1.4: Black fur color - a dominant trait - [CC BY 4.0](#)
 - 3.1.5: Yellow fur color - a recessive trait - [CC BY 4.0](#)
 - 3.1.6: Brindle color - partial dominance and epistasis - [CC BY 4.0](#)
 - 3.1.7: White spotting - when there's more than two alleles - [CC BY 4.0](#)
 - 3.1.8: KIT - embryonic lethality - [CC BY 4.0](#)
 - 3.1.9: Hemophilia - a sex-linked disorder - [CC BY 4.0](#)
 - 3.1.10: It's not all in the genes - the effect of environment - [CC BY 4.0](#)
 - 3.1.11: Pleiotropy - one gene affects more than one trait - [CC BY 4.0](#)
 - 4: Genetic Technology - *Undeclared*
 - 4.1: Biotechnology - [CC BY 4.0](#)

- [4.2: Mapping Genomes](#) - *CC BY 4.0*
- [4.3: Whole-Genome Sequencing](#) - *CC BY 4.0*
- [4.4: Genomics and Proteomics](#) - *CC BY 4.0*
 - [4.4.1: Biotechnology](#) - *CC BY 4.0*
- [Back Matter](#) - *Undeclared*
 - [Index](#) - *Undeclared*
 - [Glossary](#) - *Undeclared*
 - [Detailed Licensing](#) - *Undeclared*