The Molecular Basis of Inheritance

8

- → THE SEARCH FOR INHERITABLE MATERIAL
- → STRUCTURE OF NUCLEIC ACIDS
- → DNA REPLICATION IN EUKARYOTES
- → FROM DNA TO PROTEIN
- → GENE MUTATION
- → THE GENETICS OF VIRUSES AND BACTERIA
- → PRIONS
- → THE HUMAN GENOME
- → RECOMBINANT DNA; CLONING GENES
- → TOOLS AND TECHNIQUES OF RECOMBINANT DNA
- → ETHICAL CONSIDERATIONS

INTRODUCTION

Today, everyone knows that DNA is the molecule of heredity. We know that DNA makes up chromosomes and that genes are located on the chromosomes. Today, we can even see the location of particular genes by tagging them with fluorescent dye.

However, until the 1940s, many scientists believed that proteins, not DNA, were the molecules that make up genes and constitute inherited material. Several factors contributed to that belief. First, proteins are a major component of all cells. Second, they are complex macromolecules that exist in seemingly limitless variety and have great specificity of function. Third, a great deal was known about the structure of proteins and very little was known about DNA. The work of many brilliant scientists has transformed our knowledge of the structure and function of the DNA molecule and led to the acceptance of DNA as the molecule responsible for heredity.

This chapter includes the history of the search for the heritable material, the structure of nucleic acids and how DNA makes proteins. It also includes an extensive review of genetic engineering and recombinant DNA techniques.

THE SEARCH FOR INHERITABLE MATERIAL

Griffith (1927) performed experiments with several different strains of the bacterium *Diplococcus pneumoniae*. Some strains are virulent and cause pneumonia in humans and mice, and some strains are harmless. Griffith discovered that *bacteria have the ability to transform harmless cells into virulent ones by transferring some genetic factor from one bacteria cell to*

another. This phenomenon is known as **bacterial transformation**, and the experiment is known as the **transformation experiment**. See the information about bacterial transformation later in this chapter.

Avery, MacLeod, and McCarty (1944) published their classic findings that Griffith's **transformation factor** is, in fact, DNA. This research proved that DNA was the agent that carried the genetic characteristics from the virulent dead bacteria to the living nonvirulent bacteria. *This provided direct experimental evidence that DNA, not protein, was the genetic material.*

Hershey and Chase (1952) carried out experiments that lent strong support to *the theory that DNA is the genetic material.* They tagged bacteriophages with the radioactive isotopes ^{32}P and ^{35}S. Since proteins contain sulfur but not phosphorus and DNA contains phosphorus but not sulfur, the radioactive ^{32}P labeled the DNA of the phage viruses while ^{35}S labeled the protein coat of the phage viruses. Hershey and Chase found that when bacteria were infected with phage viruses, the radioactive phosphorus in the phage always entered the bacterium while the radioactive sulfur remained outside the cells. This proved that *DNA from the viral nucleus, not protein from the viral coat, was infecting bacteria and producing thousands of progeny.*

Rosalind Franklin (1950–53), while working in the lab of Maurice Wilkins, carried out the X-ray crystallography analysis of DNA that showed DNA to be a helix. Her work was critical to Watson and Crick. Although Maurice Wilkins shared the Nobel prize with Watson and Crick, Rosalind Franklin did not. She had died by the time the prize was awarded, and the prize is not awarded posthumously.

Watson and Crick (1953), while working at Cambridge University, *proposed the double helix structure of DNA* in a one-page paper in the British journal *Nature.* Throughout the 1940s, until 1953, many scientists worked to understand the structure of DNA. All the data that Watson and Crick used to build their model of DNA derived from other scientists who published earlier. Two major pieces of information they used were the biochemical analysis of DNA (from Erwin Chargaff) and the X-ray diffraction analysis of DNA (from Rosalind Franklin). However, the fact that much of the components of DNA were known before Watson and Crick began their model building does not detract from the brilliance of their achievement. Understanding the structure of DNA gives a foundation to understand how DNA could replicate itself. Watson and Crick received the Nobel prize in 1962 for correctly describing the structure of DNA.

Meselson and Stahl (1958) *proved that DNA replicates in a semiconservative fashion,* as Francis Crick predicted. They cultured bacteria in a medium containing heavy nitrogen (^{15}N), allowing the bacteria to incorporate this heavy nitrogen into their DNA as they replicated and divided. These bacteria were then transferred to a medium containing light nitrogen (^{14}N) and allowed to replicate and divide only once. The bacteria that resulted from this final replication were spun in a centrifuge and found to be midway in density between the bacteria grown in heavy nitrogen and those grown in light nitrogen. This demonstrated that the new bacteria contained DNA consisting of one heavy strand and one light strand. See Figure 8.1 of semiconservative replication.

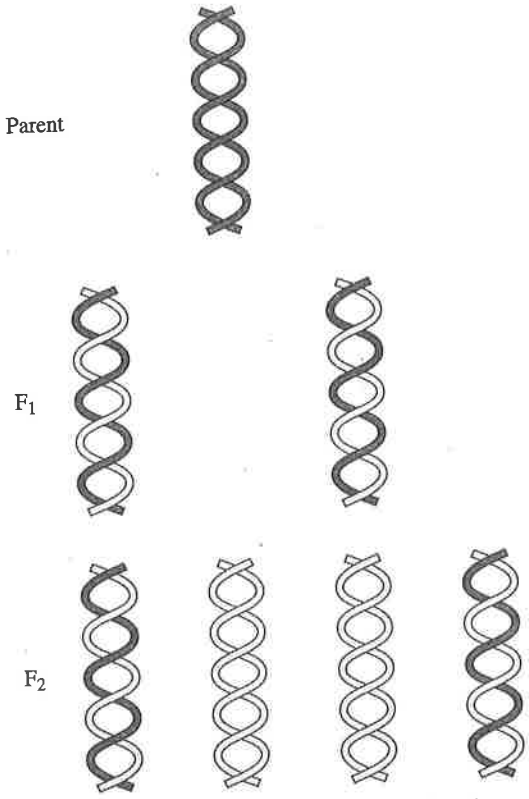


Figure 8.1 Semiconservative Replication

STRUCTURE OF NUCLEIC ACIDS

Deoxyribonucleic Acid (DNA)

The DNA molecule is a **double helix**, shaped like a twisted ladder, consisting of two strands running in opposite directions (antiparallel); see Figure 8.2. One strand runs **5′ to 3′** (right side up), the other **3′ to 5′** (upside down). DNA is a polymer consisting of repeating units of **nucleotides**. In DNA, these consist of a **5-carbon sugar** (**deoxyribose**), a **phosphate**, and a **nitrogen base**. The carbon atoms in deoxyribose are numbered 1 to 5. There are four nitrogenous bases in DNA: **adenine (A)**, **thymine (T)**, **cytosine (C)**, and **guanine (G)**. Of the four nitrogenous bases, adenine and guanine are **purines**, and thymine and cytosine are **pyrimidines**. The nitrogenous bases of opposite chains are paired to one another by **hydrogen bonds**: the adenine nucleotide bonds by a **double hydrogen bond** to the thymine nucleotide, and the cytosine nucleotide bonds by a **triple hydrogen bond** to the guanine nucleotide.

REMEMBER

The two strands of DNA run in opposite directions.

DNA gets packed and unpacked in the nucleus as needed. Eukaryotic DNA combines with a large amount of proteins called **histones** from which it separates only briefly during replication. This complex of DNA plus histones is called by the general name **chromatin**. The double helix of DNA wraps twice around a core of histones, forming structures called **nucleosomes** that look like beads on a string.

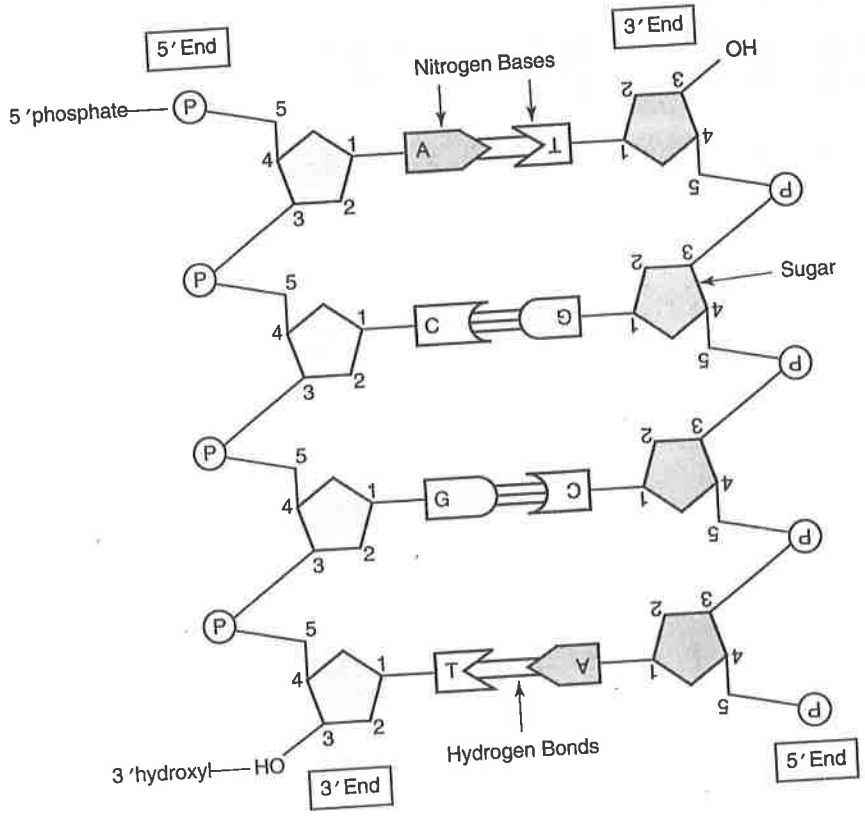


Figure 8.2 DNA

REMEMBER

A bonds with T.

A = T

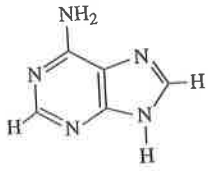
C bonds with G.

C ≡ G

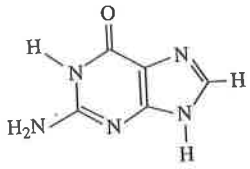
Ribonucleic Acid (RNA)

RNA is a single-stranded helix consisting of repeating nucleotides: adenine, cytosine, guanine, and **uracil (U)**, which replaces thymine. The 5-carbon sugar in RNA is **ribose**.

Figure 8.3 shows structural formulas for the purines—adenine and guanine—and for the pyrimidines—cytosine, thymine, and uracil.

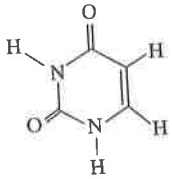


adenine (A)

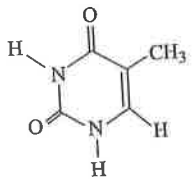


guanine (G)

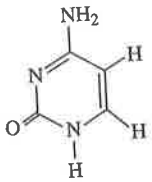
Purine Nitrogen Bases



uracil (U)



thymine (T)



cytosine (C)

Pyrimidine Nitrogen Bases

Figure 8.3

DNA REPLICATION IN EUKARYOTES

DNA replication, the making of an exact replica of the DNA molecule by **semiconservative replication**, was predicted by Watson and Crick and proven by Meselson and Stahl. The DNA double helix unzips, and each strand serves as a **template** for the formation of a new strand composed of complementary nucleotides: A with T and C with G. The two new molecules each consist of one old strand and one new strand. The following describes DNA replication in eukaryotes. See also Figures 8.1 and 8.4.

- Replication begins at special sites called **origins of replication**, where the two strands of DNA separate to form **replication bubbles**. Thousands of these bubbles can be seen along the DNA molecule by using electron microscopy. Replication bubbles speed up the process of replication along the giant DNA molecule that consists of *6 billion nucleotides*. A replication bubble expands as replication proceeds in *both directions at once.*
- At each end of the replication bubble is a **replication fork**, a Y-shaped region where the new strands of DNA are elongating. Eventually, all the replication bubbles fuse.
- The enzyme **DNA polymerase** catalyzes the antiparallel elongation of the new DNA strands. (At least 15 different types of DNA polymerase have been identified, but only one is involved in the elongation of the DNA strand.)
- DNA polymerase builds a new strand from the 5′ to the 3′ direction by moving along the template strand and pushing the replication fork ahead of it. In humans, the rate of elongation is about 50 nucleotides per second.
- *DNA polymerase cannot initiate synthesis*; it can only add nucleotides to the 3′ end of a preexisting chain. This preexisting chain actually consists of RNA and is called **RNA primer**. An enzyme called **primase** makes the primer by joining together RNA nucleotides.
- One of your cells can replicate its entire DNA in a few hours.
- DNA polymerase replicates the two original strands of DNA differently. Although it builds both new strands in the 5′ to 3′ direction, one strand is formed *toward the replication fork* in an unbroken, linear fashion. This is called the **leading strand**. The other strand, the **lagging strand**, forms in the direction *away from the replication fork* in a series of segments called **Okazaki fragments**. Okazaki fragments are about 100–200 nucleotides long and will be joined into one continuous strand by the enzyme **DNA ligase.**
- Other proteins and enzymes assist in replication of the DNA. **Helicases** are enzymes that untwist the double helix at the replication fork. They separate the two parental strands, making these strands available as templates. **Single-stranded binding proteins** act as scaffolding, holding the two DNA strands apart. **Topoisomerases** lessen the tension on the tightly wound helix by breaking, swiveling, and rejoining the DNA strands.
- DNA polymerases carry out **mismatch repair**, a kind of proofreading that corrects errors. Damaged regions of DNA are excised by **DNA nuclease.**
- Each time the DNA replicates, some nucleotides from the ends of the chromosomes are lost. To protect against the possible loss of genes at the ends of the chromosomes, eukaryotes have special nonsense nucleotide sequences (TTAGGG) at the ends of the chromosomes that repeat thousands of times. These protective ends are called **telomeres**. Telomeres are created and maintained by the enzyme **telomerase**. Normal body cells contain little telomerase, so every time the DNA replicates, the telomeres get shorter. This may serve as a clock that counts cell divisions and causes the cell to stop dividing as the cell ages.

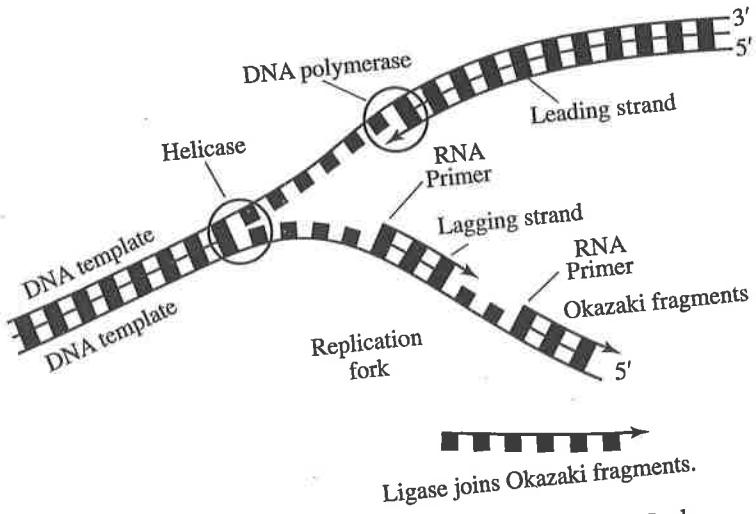


Figure 8.4 DNA Replication at Replication Fork

FROM DNA TO PROTEIN

The process whereby DNA makes proteins has been worked out in great detail. To summarize, the **triplet code** in DNA is **transcribed** into a **codon sequence** in messenger-RNA (mRNA) inside the nucleus. Next, this newly formed strand of **RNA**, known as pre-RNA, is **processed** or modified in the nucleus. Then the codon sequence leaves the nucleus and is **translated** into an amino acid sequence (a polypeptide) in the cytoplasm at the ribosome.

If the strand of DNA triplets to be transcribed is 5′-AAA TAA CCG GAC-3′

Then the strand of mRNA **codons** that forms is 3′-UUU AUU GGC CUG-5′

The transfer RNA (tRNA) **anticodon** strand complementary to the mRNA strand is AAA UAA CCG GAC

Figure 8.5 shows an overview of transcription, RNA processing, and translation.

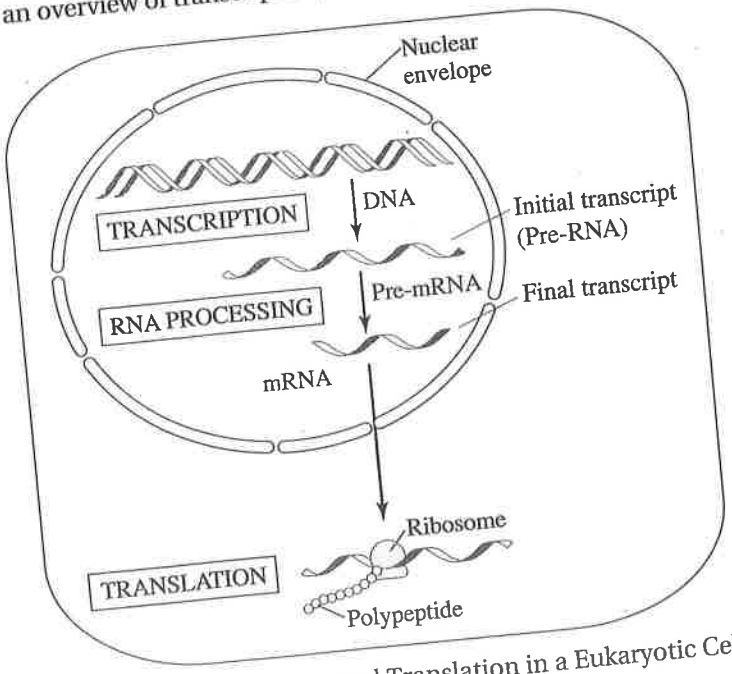


Figure 8.5 Transcription and Translation in a Eukaryotic Cell

Transcription

Transcription is the process by which the information in a DNA sequence is copied (transcribed) into a complementary RNA sequence. Although there are many kinds of RNA, three types are directly involved in protein synthesis. See Figure 8.6.

1. **MESSENGER RNA (mRNA) IS INVOLVED IN TRANSCRIPTION:** When a sequence of DNA is expressed, one of two strands of DNA is copied into mRNA according to the base-pairing rules, C with G and A with U (in RNA, uracil replaces the thymine in DNA).
2. **RIBOSOMAL RNA (rRNA) IS INVOLVED IN TRANSLATION:** rRNA is structural. Along with proteins, it makes up the ribosome, which consists of two subunits, one large and one small. The ribosome has one mRNA binding site and three tRNA binding sites, known as A, P, and E sites. A ribosome is a protein synthesis factory.
3. **TRANSCRIPTION RNA (tRNA) CARRIES AMINO ACIDS FROM THE CYTOPLASMIC POOL OF AMINO ACIDS TO mRNA AT THE RIBOSOME:** tRNA is shaped like a coverleaf and has a binding site for an amino acid at one end and another binding site for an **anticodon** sequence that binds to mRNA at the other.

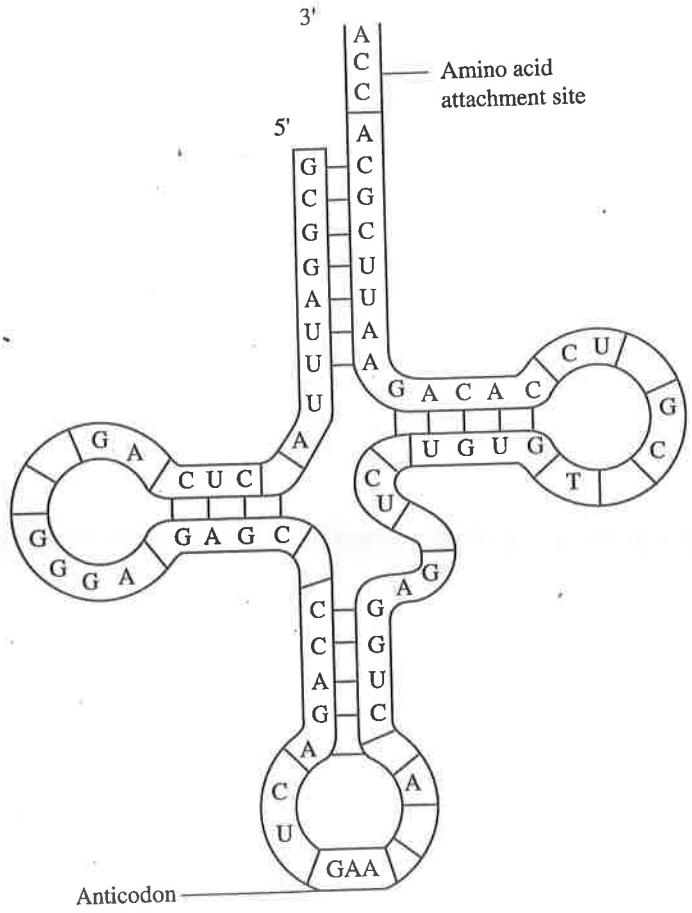


Figure 8.6 Transfer-RNA (tRNA)

- Transcription consists of three stages: **initiation**, **elongation**, and **termination**.
- **Initiation** begins when an enzyme, ***RNA polymerase***, *recognizes and binds to DNA at the* ***promoter region***. The promoter "tells" RNA polymerase where to begin transcription and which of the two strands to transcribe. A collection of proteins called **transcription factors** recognize a key area within the promoter, the **TATA box** (because of its many repeating thymine and adenine nucleotides), and mediate the binding of RNA polymerase to the DNA. The completed assembly of transcription factors and RNA polymerase bound to the promoter is called a **transcription initiation complex**. Once RNA polymerase is attached to the promoter, DNA transcription of the DNA **template** begins.
- **Elongation** of the strand continues as *RNA polymerase adds nucleotides to the 3′ end of a growing chain*. RNA polymerase pries the two strands of DNA apart and attaches RNA nucleotides according to the base pairing rules: C with G and A with U. The stretch of DNA that is transcribed into an mRNA molecule is called a **transcription unit**. Each unit consists of triplets of bases called **codons** (for example, AAU, CGA) that code for specific amino acids. A single gene can be transcribed into mRNA simultaneously by several molecules of RNA polymerase following each other in a caravan fashion. Like DNA polymerases, RNA polumerase has mechanisms for proofreading during transcription. Because mRNA is usually short-lived, any errors in mRNA are not as potentially harmful as errors in the DNA sequence.
- **Termination** is the final stage in transcription. Elongation continues for a short distance after the RNA polymerase transcribes the **termination sequence** (AAUAAA). At this point, mRNA is cut free from the DNA template.

RNA Processing

Before the newly formed pre-RNA strand is shipped out of the nucleus to the ribosome in the cytoplasm, it is altered or **processed** by a series of enzymes. Here are the details.

- A **5′ cap** consisting of a modified guanine nucleotide is added to the 5′ end. This cap helps the RNA strand bind to the ribosome in the cytoplasm during translation.
- A **poly (A) tail**, consisting of a string of adenine nucleotides, is added to the 3′ end. This tail protects the RNA strand from degradation by hydrolytic enzymes, and facilitates the release of mRNA from the nucleus into the cytoplasm.
- Noncoding regions of the mRNA called **introns** or **intervening sequences** are removed by **snRNPs**, small nuclear ribonucleoproteins, and **splicesomes**. This removal allows only **exons**, which are expressed sequences, to leave the nucleus. As a result of this processing, the mRNA that leaves the nucleus is a great deal shorter than the original transcription unit. See Figure 8.5 on page 134.

Alternative Splicing

Before the human genome was sequenced by the Human Genome Project, scientists expected that they would find about 100,000 genes. In fact, they discovered that humans have only about 24,000 genes. This surprised everyone but can be explained when you recognize that different mRNAs can be synthesized from the same primary transcript. In **alternative RNA splicing**, different RNA molecules are produced from the same primary transcript, depending on which RNA segments are treated as exons and which as introns. *Regulatory proteins* specific to a cell type control intron-exon choices by binding to regulatory sequences within the primary transcript. See Figure 8.7.

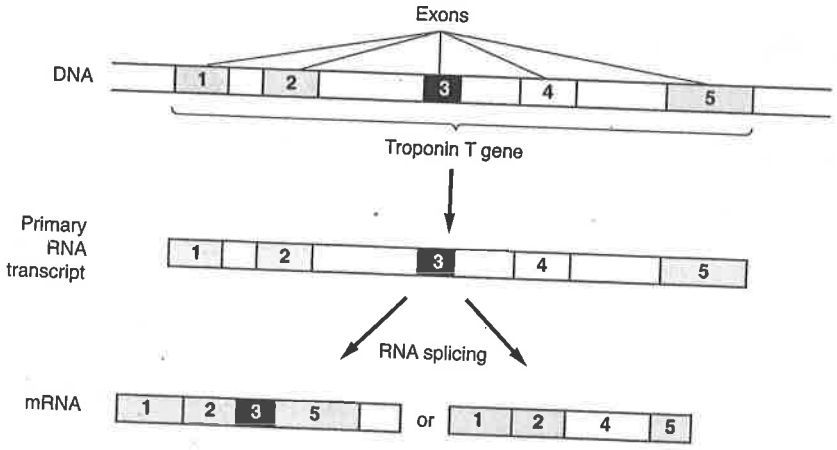


Figure 8.7 Alternative RNA Splicing

Translation of mRNA—Synthesis of a Polypeptide

Translation *is the process by which the codons of an mRNA sequence are changed into an amino acid sequence*; see Figure 8.8. Amino acids present in the cytoplasm are carried by tRNA molecules to the codons of the mRNA strand at the ribosome according to the base pairing rules (A with U and C with G). One end of the tRNA molecule bears a specific amino acid, and the other end bears a nucleotide triplet called an **anticodon**. Unlike mRNA which is broken down immediately after it is used, tRNA is used repeatedly. The energy for this process is provided by **GTP (guanosine triphosphate)**, a molecule closely related to ATP. Each amino acid is joined to the correct tRNA by a specific enzyme called **aminoacyl-tRNA synthetase**. There are only 20 different aminoacyl-tRNA synthetases, one for each amino acid. There are 64 codons; 61 of them code for amino acids. One codon, **AUG**, has two functions; it codes for methionine and is also a **start codon**. Three codons, **UAA**, **UGA**, and **UAG**, are **stop codons**, and terminate all sequences. Some tRNA molecules have anticodons that can recognize two or more different codons. This occurs because the pairing rules for the third base of a codon are not as strict as they are for the first two bases. This relaxation of base pairing rules is known as **wobble**. For example, the codons UCU, UCC, UCA, and UCG all code for the amino acid serine.

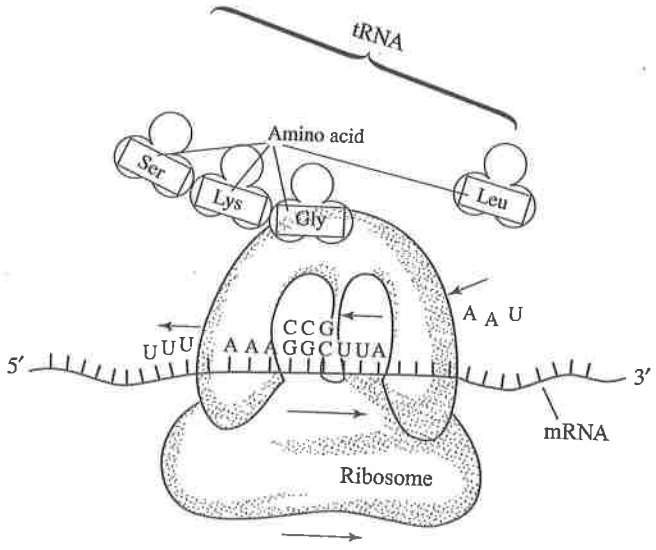


Figure 8.8 Translation

See Figure 8.6 for a sketch of tRNA and Figure 8.8 for a sketch of translation. The process of translation consists of three stages: **initiation**, **elongation**, and **termination**.

- **Initiation** begins when mRNA becomes attached to a subunit of the ribosome. This first codon is always **AUG**. It must be positioned correctly in order for transcription of an amino acid sequence to begin.
- **Elongation** continues as tRNA brings amino acids to the ribosome and a polypeptide chain is formed. One mRNA molecule is generally translated simultaneously by several ribosomes in clusters called **polyribosomes**.
- **Termination** of an mRNA strand is complete when a ribosome reaches one of three **termination** or **stop codons**. A **release factor** breaks the bond between the tRNA and the last amino acid of the polypeptide chain. The polypeptide is freed from the ribosome, and mRNA is broken down.

The Genetic Code

The complete genetic code is shown in Figure 8.9. There are 64 (4^3) possible combinations of the four nitrogenous bases. Notice that AUG, which codes for methionine, also codes for **start**, the initiation signal for translation. Three of the codons are **stop codons**, termination signals of translation.

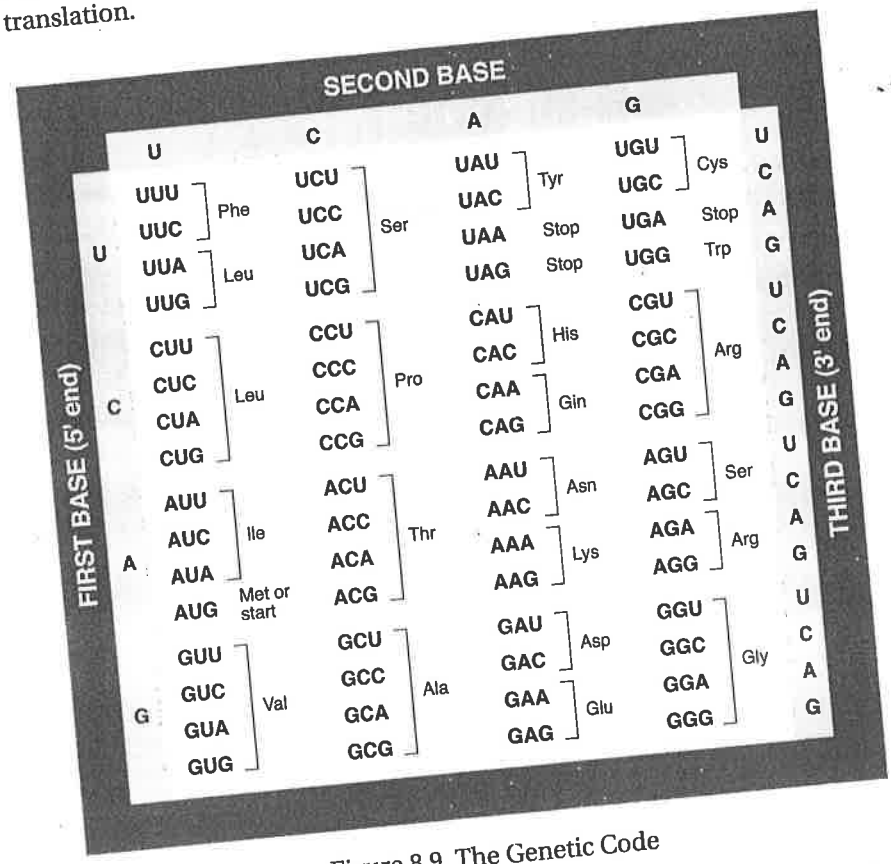


Figure 8.9 The Genetic Code

Remember this important statement about the genetic code: *There are redundancies in the code, but there is no ambiguity.* Examine the chart. The fact that there are four codons for leucine means there is redundancy. In every case, one codon codes for one particular amino acid. So there is no ambiguity.

With very few exceptions within bacteria, mitochondria and chloroplasts, this code is universal and unifies all life. It indicates that the code originated early in the evolution of life on Earth and that all living things descended from those first ancestral cells.

GENE MUTATION

Mutations are permanent changes in genetic material. They occur *spontaneously* and *randomly*. They can be caused by **mutagenic agents**, including toxic chemicals and radiation. A mutation in **somatic** (body) cells disrupts normal cell functions. Mutations that occur in gametes are transmitted to offspring and can change the **gene pool** of a population. *Mutations are the raw material for natural selection.*

Some regions of DNA are more vulnerable to mutations than others. For example, regions of As and Ts are subject to more breakages than regions of Cs and Gs because A and T are connected by a double hydrogen bond whereas Cs and Gs are connected by a triple hydrogen bond.

Point Mutation

The simplest mutation is a **point mutation**. This is a **base-pair substitution**, a chemical change in just one base pair in a single gene. Here is an example of a change in an English sentence analogous to a point mutation in DNA:

Point Mutation

THE FAT CAT SAW THE **D**OG → THE FAT CAT SAW THE **H**OG.

The inherited genetic disorder **sickle cell anemia** results from a single point mutation in a single base pair in the gene that codes for hemoglobin. This point mutation is responsible for the production of abnormal hemoglobin that can cause red blood cells to sickle when oxygen tension is low. When red blood cells sickle, a variety of tissues may be deprived of oxygen and suffer severe, permanent damage. The possibility exists, however, that a point mutation could result in a beneficial change for an organism or, because of **wobble** in the genetic code, result in no change in the proteins produced. (Wobble is the relaxation of the base-pairing rules for the third base in a codon.) Here is an example of wobble:

DNA	mRNA	Amino Acid Produced
AAA	UUU	Phenylalanine
AAG	UUC	Phenylalanine
↑ mutation		No change occurs in the amino acid.

Insertion or Deletion

A second type of gene mutation results from a single nucleotide **insertion** or **deletion**. To continue the three-letter word analogy, a deletion is the loss of one letter, and an insertion is the addition of a letter into the DNA sentence. Both mutations result in a **frameshift**, because the entire reading frame is altered.

Deletion of the Letter E shifts the reading frame.

THE FAT CAT SAW THE DOG → THF ATC ATS AWT HED OG

Insertion of the Letter T shifts the reading frame.

THE FAT CAT SAW THE DOG → THE FTA TCA TSA WTH EDO G

As a result of the frameshift, one of two things can happen. Either a mutated polypeptide is formed or no polypeptide is formed.

Missense Mutations

When point mutations or frameshifts change a codon *within a gene* into a stop codon, translation will be altered into a **missense** or **nonsense mutation.**

THE GENETICS OF VIRUSES AND BACTERIA

Since the early part of the twentieth century when Griffith discovered the transformation factor, knowledge of genetics has been based on work with the simplest biological systems—viruses and bacteria. Scientists' understanding of replication, transcription, and translation of DNA was worked out using bacteria as a model. Their understanding of how viruses and bacteria infect cells is the basis for how diseases are treated and how vaccines are developed. A worldwide industry of genetic engineering and recombinant DNA relies on bacteria like *Escherichia coli* and viruses like the phage viruses for research and therapeutic endeavors. Whereas Gregor Mendel depended on the garden pea and Thomas Hunt Morgan on the fruit fly, researchers now depend on bacteria and viruses.

The Genetics of Viruses

A virus is a parasite that can live only inside another cell. It commandeers the host cell machinery to transcribe and translate all the proteins it needs to fashion new viruses. In the process, thousands of new viruses are formed and the host cell is often destroyed. A virus consists of DNA or RNA enclosed in a protein coat called a **capsid**. Some viruses also have a viral **envelope** that is derived from membranes of host cells, cloaks the capsid, and aids the virus in infecting the host. Each type of virus can infect only one specific cell type because it gains entrance into a cell by binding to *specific receptors* on the cell surface. For example, the virus that causes colds in humans infects only the membranes of the respiratory system, and the virus that causes AIDS infects only one type of white blood cell. In addition, one virus can usually only infect one species. The range of organisms that a virus can attack is referred to as the **host range** of the virus. A sudden emergence of a new viral disease that affects humans, such as AIDS or H1N1, may result from a mutation in the virus that expands its host range.

- **BACTERIOPHAGES**—The most complex and best understood virus is the one that infects bacteria, the **bacteriophage**, or **phage** virus. The bacteriophage can reproduce in different ways.

  1. In the **lytic cycle**, the phage enters a host cell, takes control of the cell machinery, replicates itself, and then causes the cell to burst, releasing a new generation of infectious phage viruses. These new viruses infect and kill thousands of cells in the same manner. A phage that replicates only by a lytic cycle is a virulent phage.
  2. In the **lysogenic cycle**, viruses replicate without destroying the host cell. The phage virus becomes incorporated into a specific site in the host's DNA. It remains dormant within the host genome and is called a **prophage**. As the host cell divides, the phage is replicated along with it and a single infected cell gives rise to a population of infected cells. At some point, an environmental trigger causes the prophage to switch to the **lytic phase**. Viruses capable of both modes of reproducing, lytic and lysogenic, within a bacterium are called **temperate viruses**.

- **RETROVIRUSES** are viruses that contain RNA instead of DNA and replicate in an unusual way. Following infection of the host cell, the retrovirus RNA serves as a template for the synthesis of complementary DNA (cDNA) because it is complementary to the RNA from which it was copied. *Thus, these retroviruses reverse the usual flow of information from DNA to RNA.* This reverse transcription occurs under the direction of an enzyme called **reverse transcriptase**. A retrovirus usually inserts itself into the host genome, becomes a permanent resident, called a **prophage**, and is capable of making multiple copies of the viral genome for years. An example of a retrovirus is the HIV (human immunodeficiency virus), which causes AIDS.
- **TRANSDUCTION**—Phage viruses acquire bits of bacterial DNA as they infect one cell after another. This process, which leads to genetic recombination, is called **transduction**. Two types of transduction occur, **generalized** and **restricted (specialized)**. Generalized transduction moves random pieces of bacterial DNA as the phage lyses one cell and infects another during the lytic cycle. **Restricted transduction** involves the transfer of specific pieces of DNA. During the lysogenic cycle, a phage integrates into the host cell at a specific site. At a later time, when the phage ruptures out of the host DNA, it sometimes carries a piece of adjacent host DNA with it and inserts this host DNA into the next host it infects.

The Genetics of Bacteria

The bacterial chromosome is a circular, double-stranded DNA molecule, tightly condensed into a structure called a **nucleoid**, which has no nuclear membrane. Bacteria replicate their DNA in **both directions** from a **single point of origin**.

Although bacteria can reproduce by a primitive sexual method called **conjugation**, the main mode of reproduction is asexual, by **binary fission**. Binary fission results in a population with all identical genes, but mutations do occur spontaneously. Although mutations are rare, bacteria reproduce by the millions, and even one mutation in every 1,000 replications can amount to significant variation in the population as a whole.

- **Bacterial transformation** was discovered by **Frederick Griffith** in 1927 when he performed experiments with several different strains of the bacterium *Diplococcus pneumoniae.*

  Transformation is either a natural or an artificial process that provides a mechanism for the recombination of genetic information in some bacteria. Small pieces of extracellular DNA are taken up by a living bacterium, ultimately leading to a stable genetic change in the recipient cell. Bacterial transformation is very easy to carry out today.
- A **plasmid** is a foreign, small, circular, self-replicating DNA molecule that inhabits a bacterium. A bacterium can harbor many plasmids and will express the genes carried by the plasmid. The first plasmid discovered was the **F plasmid**. F stands for fertility. Bacteria that contain the F plasmid are called F^+; those that do not carry the plasmid are called F^-. The F plasmid contains genes for the production of **pili**, cytoplasmic bridges that connect to an adjacent cell and that allow DNA to move from one cell to another in a form of primitive sexual reproduction called **conjugation**. Another plasmid, the **R plasmid**, makes the cell in which it is carried resistant to specific antibiotics, such as ampicillin or tetracycline. In addition, the R plasmid can be transferred to other bacteria by conjugation. Bacteria that carry the R plasmid have a distinct evolutionary advantage over bacteria that are not resistant to antibiotics. Resistant bacteria will be selected for (survive) and their populations will increase while nonresistant bacteria die out. This is exactly what is happening today as an increasing number of populations of pathogenic bacteria, such as the one that causes tuberculosis, are becoming resistant to antibiotics. This is cause for serious concern in the health community.

The Operon

REMEMBER

If an essay question on the AP exam is about regulation, the operon is a perfect example.

The **operon** was discovered in the bacterium *E. coli* by **Jacob and Monod** in the 1940s. Although it is found only in bacteria, it is an important model of **gene regulation.** An operon is essentially a set of genes and the switches that control the expression of those genes. There are two types of operons: the **inducible (lac)** operon and the **repressible (tryptophan)** operon.

THE TRYPTOPHAN OPERON

The **tryptophan operon** consists of a **promoter** and five adjacent structural genes (A, B, C, D and E) that code for the five separate enzymes necessary to synthesize the amino acid tryptophan; see Figures 8.10 and 8.11. As long as **RNA polymerase** binds to the promoter, one long strand of mRNA containing start and stop codons is transcribed. If adequate tryptophan is present, tryptophan itself acts as a **corepressor** activating the **repressor.** The activated repressor binds to the **operator**, preventing RNA polymerase from binding to the promoter. Without RNA polymerase attached to DNA at the promoter, transcription ceases. The tryptophan operon is known as a repressible operon, meaning it is always switched on unless the repressor is activated.

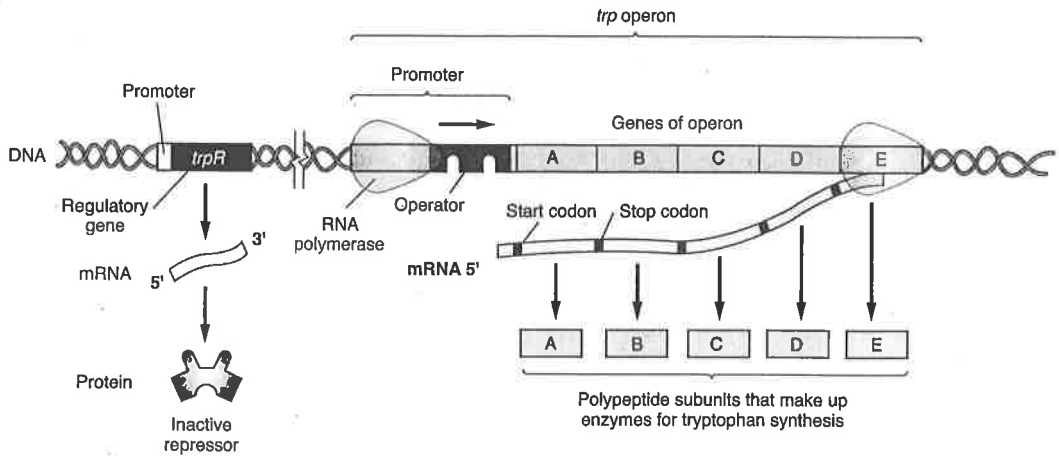


Figure 8.10 Tryptophan absent, repressor inactive, operon on, → tryptophan produced

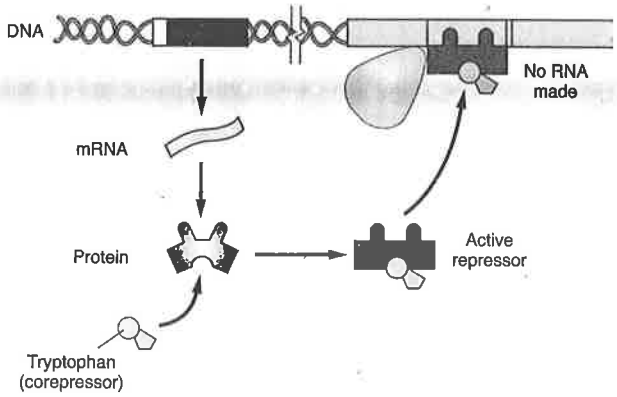


Figure 8.11 Tryptophan present, repressor active, operon off

THE LAC OPERON

In order for the *E. coli* in our intestines to utilize lactose as an energy source, three enzymes must be synthesized to break down lactose into glucose and galactose. These enzymes, β-galactosidase, permease, and transacetylase, are coded for by three genes in the *lac* operon (A, B, and C); see Figures 8.12 and 8.13. In order for these three genes to be transcribed, the **repressor** must be prevented from binding to the operator and RNA polymerase must bind to the promoter region. Allolactose, an isomer of lactose, is the **inducer** that facilitates this process by binding to the **active repressor** and inactivating it. When a person drinks milk, they ingest allolactose, the inducer, which deactivates the repressor, allowing RNA polymerase to bind to DNA. When RNA polymerase binds to DNA, transcription of the *lac* genes occurs and lactose can be utilized as an energy source.

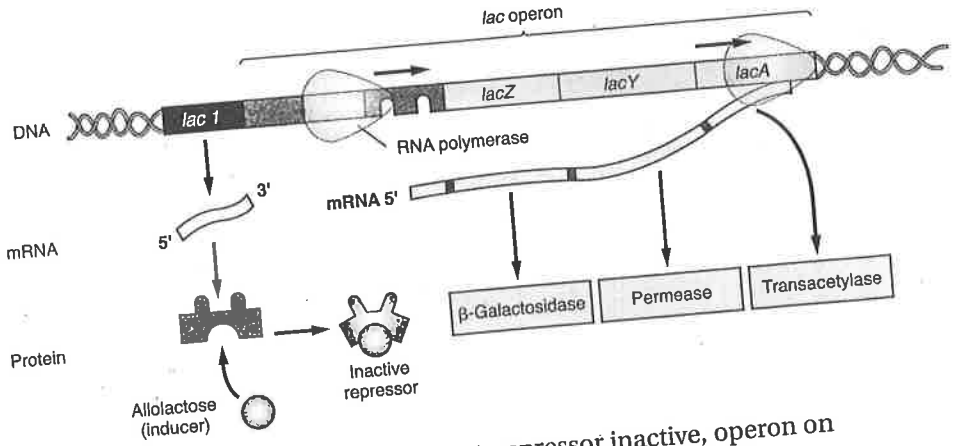


Figure 8.12 Lactose present, repressor inactive, operon on

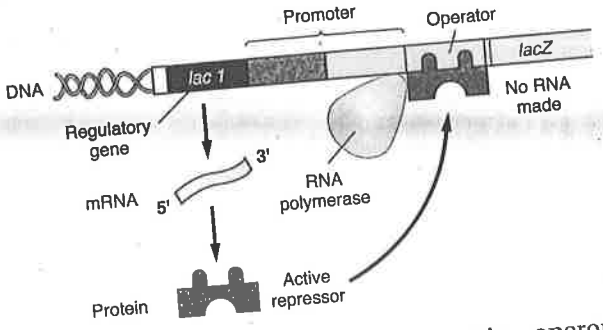


Figure 8.13 Lactose absent, repressor active, operon off

CAP AND cAMP—POSITIVE GENE REGULATION

When glucose and lactose are both present in the intestine, *E. coli* preferentially metabolize glucose and the enzymes for breaking down glucose are always present. However, when lactose is present and glucose is in short supply, *E. coli* switch to lactose as an energy source. This ability depends on the interaction of an allosteric regulatory protein, CAP (catabolite activator protein) and cAMP (cyclic AMP). Since the attachment of CAP to the promoter directly stimulates gene expression, this mechanism is an example of **positive gene regulation**.

VOCABULARY FOR THE OPERON

- **RNA polymerase**: Enzyme that transcribes a new RNA chain by linking ribonucleotides to nucleotides on a DNA template
- **Operator**: Sequence of nucleotides near the start of an operon to which the active repressor can attach. The binding of the repressor prevents RNA polymerase from attaching to the promoter and transcribing the operon's genes.
- **Promoter**: Nucleotide sequence in the DNA of a gene that is the binding site of RNA polymerase, positioning the RNA polymerase to begin to transcribe RNA at the appropriate position.
- **Repressor**: Protein that inhibits gene transcription. In the operon of prokaryotes, repressors bind to the operator.
- **Regulator gene**: Gene that codes for a repressor. It is located some distance from its operon and has its own promoter.