



UNWINDING THE DOUBLE HELIX

A GENETICS PRIMER



UNWINDING THE DOUBLE HELIX

A GENETICS PRIMER



Copyright © 1993
I.S.M., Inc.



G-NETIX

Copyright ©1993 • I.S.M., Inc.

Written by
Myo Thant, MD, MES

Graphic Design & Illustrations by
Ron Gull

Acknowledgement
Fernando E. Dela Cruz

*Additional copies of this tutorial may be purchased at \$9.00 per copy from
I.S.M., Inc., 2103 Harmony Woods Road, Owings Mills, MD 21117*

Contents

Introduction and brief history	4
The cell	9
Proteins and amino acids	17
Genes	20
How proteins are formed	33
DNA replication	44
Gene control and development	49
Mutation and recombination	61
Chromosomes	66
Mendel's Findings	71
Before a baby is born	79

Chapter 1

Introduction and Brief History of Genetics

Introduction

There are many articles written about breakthroughs in the science of genetics, either in diagnosis and treatment of diseases, breeding of animals, or breeding of plants. One new advance is **genetic engineering**. This is a method where scientists use experimental techniques to copy and produce new genes or new combinations of genes. Scientists can now isolate almost any gene from plants and animals, make copies of it and incorporate it into viruses, bacteria, and other species. They can also reproduce the gene in large quantities. It can then be inserted into animals and plants to produce new breeds, to correct a defective gene, or to produce large quantities of hormones and enzymes.

Another advance in the science of genetics is the attempt to cure a disease by gene transplant therapy, where a normal gene is inserted into the diseased cell and implanted into a patient.

The knowledge of genetics is also applied in the fields of agriculture where genetically engineered plants and fruits are grown for better and healthier products.

In health and medicine advances in genetics have helped identify and isolate gene defects in certain diseases. For example, scientists have found that certain genes called **oncogenes** are associated with many cancers.

The knowledge of genetics is also used in criminal law. Genetic sequences are used to identify individuals. Such sequences, called **DNA fingerprints**, are said to be more accurate than regular fingerprints in establishing the identity of a person. Because there are millions of genetic sequences, it is extremely unlikely that any two individuals will have the same sequence, i.e. DNA prints.

As we explore advances in genetics, the greatest and largest genetic laboratory is the planet Earth itself. Many species have evolved by mutations and adaptations. It is unfortunate that human beings are changing this environment through pollution and the destruction of nature, destroying many species that have taken years to evolve. In doing so, we are destroying the perfect genetic evolution which allows plants and animals to live in symbiotic harmony.

What is genetics? Genetics is the branch of biology that studies how living things, such as humans, animals, and plants, pass information to their descendants. It deals with heredity and variation. It studies genes which are the blueprints of life.

This tutorial is not a textbook on genetics. There are many excellent books that discuss genetics in detail. Some of them are listed in the reference section. This

tutorial will introduce the user to the basic principles which define genes and their functions. The user needs to understand the contents of this tutorial to be able to understand and play G-NETIX.

A chapter on embryology is also included. Embryology is a study of how a baby develops in the mother's womb. Again the tutorial does not intend to teach a subject as complicated as embryology. It just briefly touches on the development of a human baby as it relates to the game.

A brief history in the evolution of the science of genetics

1866: Gregor Johann Mendel (1822-1884), an Austrian monk, experimented with garden peas in a small monastery garden and published his findings.

1883: The concept of chromosomes in the nucleus of the cell was postulated by a scientist named Wilhelm Roux.

1902: Evidence that a gene is part of a chromosome was confirmed.

1905: The name "genetics" was given to this developing science. The term is derived from a Greek word meaning "to generate". The word "allele" is given to identify members of chromosome pairs which give different alternative characteristics.

1909: The word "gene" was introduced by a Danish scientist, W.L. Johannsen.

1952: DNA was demonstrated to be the genetic material .

1953: James Watson and Francis Crick demonstrated that DNA has a double-helix structure. They won Nobel prize in 1962 for their work.

1956: The normal chromosome number in a human being is found to be 46.

1961: Two scientists named Francois Jacob and Jacques Monod proposed the "operon model" for regulating gene expression. Both received the Nobel prize in 1965 for their work.

1966: The complete genetic code was determined by Scientists Marshall Nirenberg and H. Gobind Khorana. They received a Nobel prize for their work in 1968.

1972: A recombinant DNA was produced in the lab by by a scientist Walter Berg. He won the Nobel prize in 1980 for his work.

1977: Introns were first demonstrated. Introns are segments of genes that do not provide biological information.

1989: NIH Recombinant DNA Advisory Committee recommends approval of first human gene

transplant experiment.

1989: A clone was made of a gene for a disease called cystic fibrosis.

1992: Gene transfer therapy was started in patients with certain cancers and genetic deficiency diseases.

1992: Human Genome Project, an ambitious attempt to identify all human genes, is started.

Some terminology used in genetic engineering:

Chromosome map: A map showing locations of genes on the chromosomes.

DNA cloning: The technique of producing identical copy of a DNA fragment.

DNA sequencing: The technique of determining nucleotide sequence of a DNA fragment.

Genetic engineering: The technique of manipulating genes or producing new combinations of genes.

Gene splicing: The process of removing intron sequences and reconnecting exon sequences of RNA.

Recombinant DNA: A DNA molecule produced by genetic engineering.

Chapter 2

The Cell

The human body, like other animals and plants, consists of millions of cells. These cells have different appearances and different functions. For example, muscle cells are long with many nuclei and can be stretched like elastic. Gland cells, on the other hand, produce secretions and juices which are used for various functions. It is amazing to realize all these different cells, each with a specialized function, start from a single cell. This differentiation is predetermined by genes.

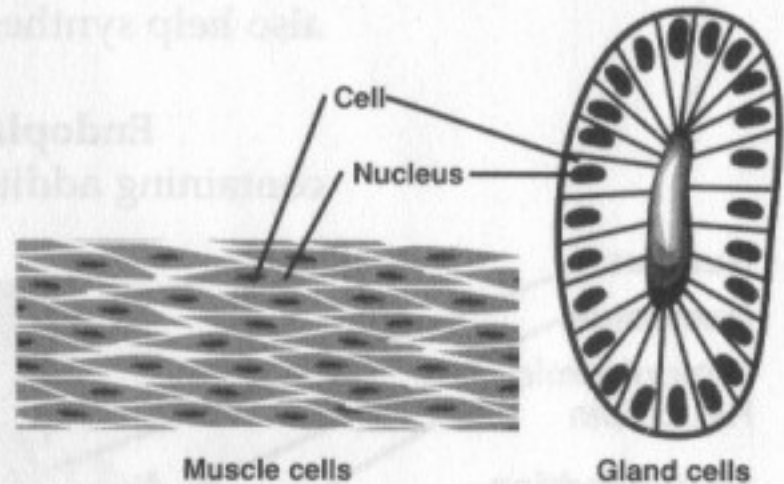


Figure 2.1. Types of cells.

Cell Structure

A human cell has a **cytoplasm** and a **nucleus**.

The **cytoplasm** is enclosed by a membrane which forms the outer boundary of the cell. Cytoplasm contains firm supporting proteins, called **cytoskeleton**, and thousands of biochemical molecules formed from elements such as water, oxygen, carbon, nitrogen, and sodium.

In addition, it contains numerous small structures called **organelles** (meaning little organs) such

G-NETIX

as mitochondria, endoplasmic reticulum, golgi vesicles and lysosomes. Organelles are surrounded by their own membranes. These organelles perform specific functions.

Mitochondria contain **enzymes**. Enzymes are proteins that speed up chemical reactions. Enzymes also help synthesize other proteins.

Endoplasmic reticulum are tubes of membrane containing additional membranes and **ribosomes** that

synthesize proteins. Endoplasmic reticulum are also involved in synthesis and breakdown of fat.

Ribosomes are complex structures that contain ribosomal RNAs (we will learn more about rRNAs in the chapter on genes) and proteins.

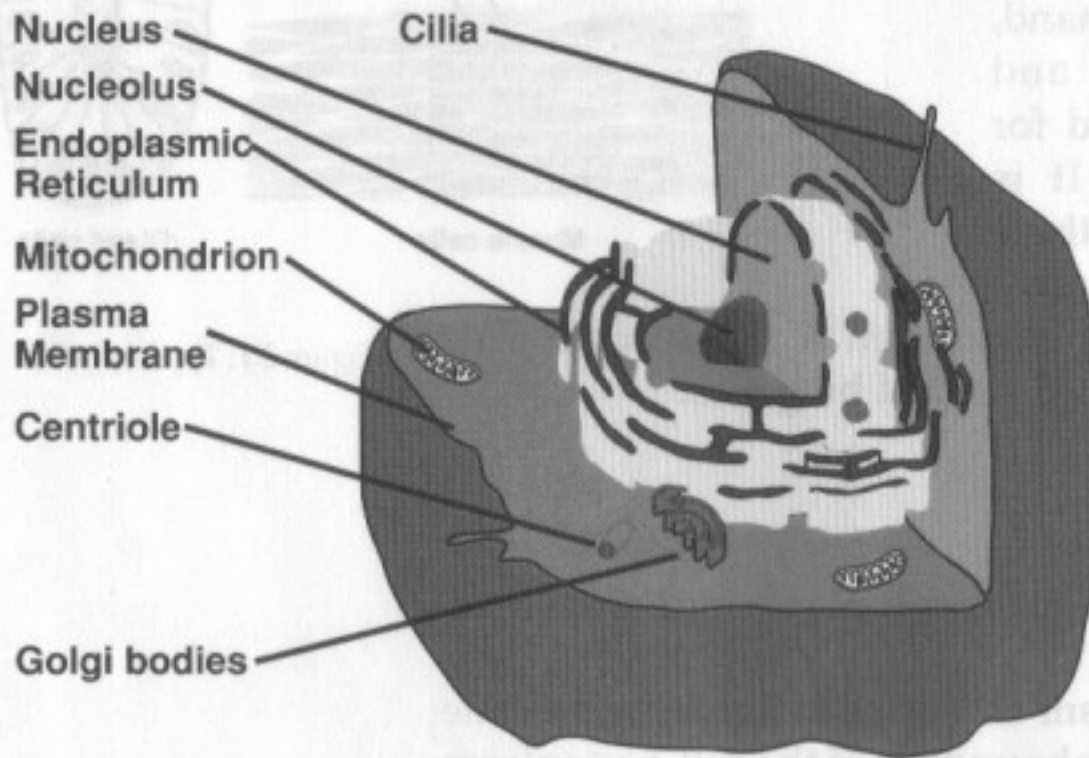


Figure 2.2. Cell structure.

Golgi vesicles are sacs that are involved in the packaging of molecules. They are needed for secretion or delivery to other parts of the cell.

Lysosomes contain enzymes that break down proteins, fats and nucleic acids.

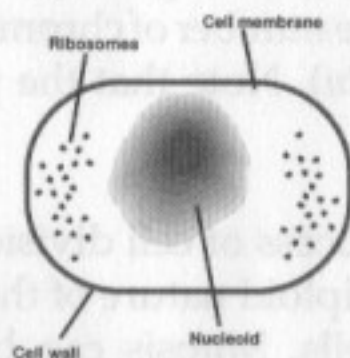
Inside the cell is a nucleus. The **nucleus** is separated from the cytoplasm by a thin membrane called **nuclear membrane**. Inside the nucleus are the **chromosomes**. Chromosomes contain **genes**, which are the essence of life. It has been calculated that a human cell contains between 30,000 and 40,000 different genes.

Also inside the nucleus is a **nucleolus**, where the ribosomal RNAs are synthesized. Therefore, ribosomal RNAs are synthesized in the nucleolus.

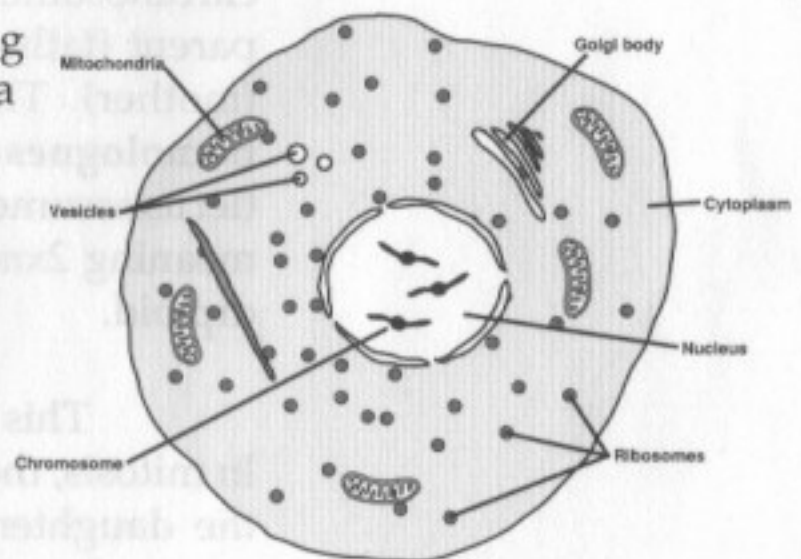
Eukaryote and Prokaryote

Eukaryotes and **prokaryotes** are two types of cells which have existed for more than a billion of years.

A **Prokaryote** is an organism consisting of one cell. Its structure is simple. It has a cell wall. The organelles of prokaryote have no membranes. There is no true nucleus. The genes are just clumped together. A bacteria is an example of a prokaryote.



Prokaryote



Eukaryote

Figure 2.3. Prokaryote and Eukaryote.

A **Eukaryote**, on the other hand, may consist of one

or more cells. It has many intracellular organelles, such as mitochondria, and a nucleus that contains chromosomes and genes. Human beings, animals, and plants are examples of eukaryotes.

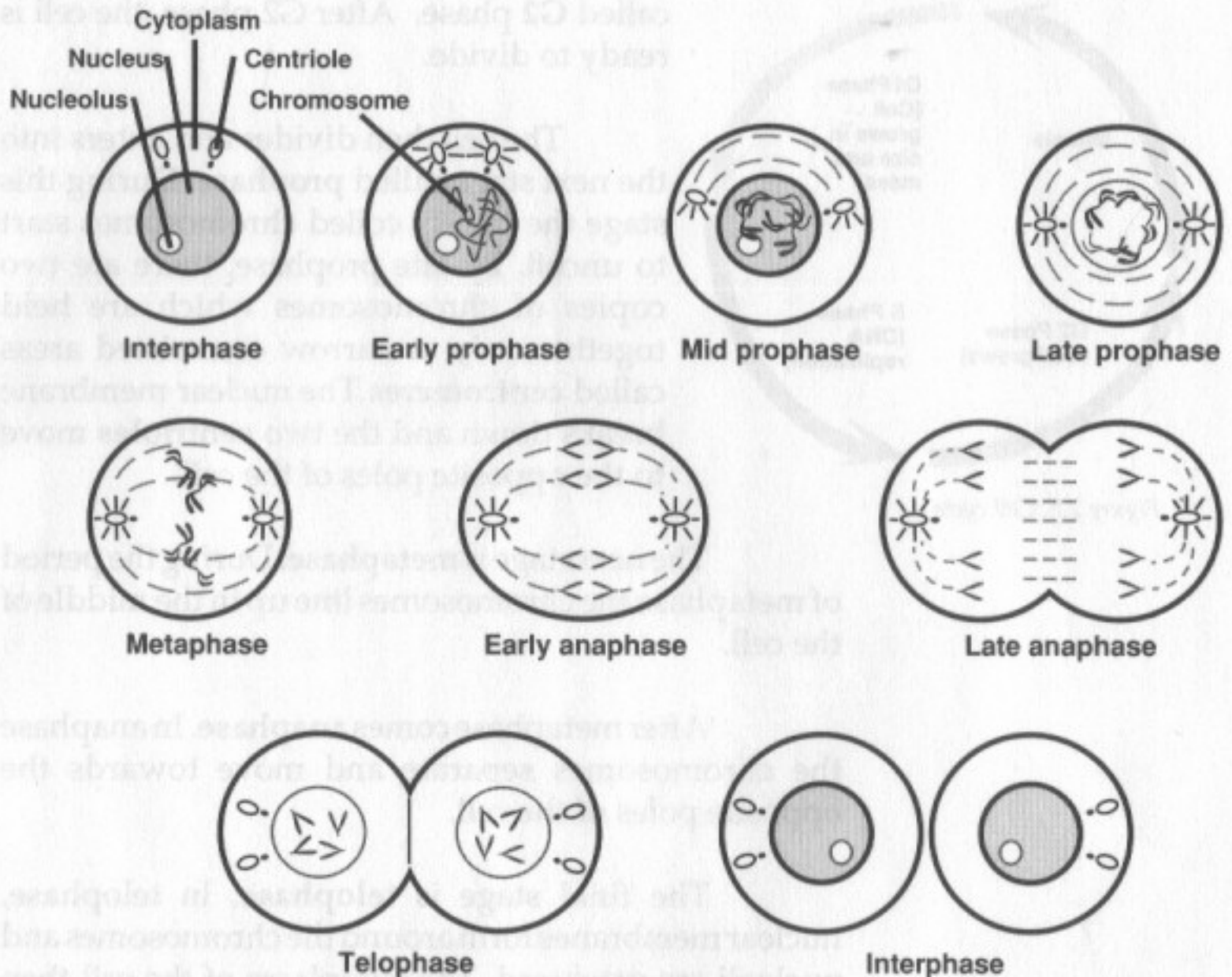
It is necessary to understand the difference between eukaryote and prokaryote since their genetic structures, functions and expressions are each distinct.

Cell Division and Cell Cycle

The embryo starts with one cell called the **zygote**. This cell divides into two daughter cells, then four cells, then eight cells and so on. By the time the embryo has grown into an adult there are millions of cells in the body.

Each daughter cell, except for the sperm and egg, has two copies of each morphologically distinct chromosome. One copy is inherited from the male parent (father) and one copy from the female parent (mother). The two paired chromosomes are called **homologues** and the daughter cell is said to be **diploid** (let us assume the number of chromosomes is n . Diploid, meaning $2 \times n = 2n$). Note that the parent cells are also diploid.

This process of cell division is called **Mitosis**. In mitosis, the diploid nature of the cell is passed on to the daughter cells. Mitosis can be divided into five stages: **interphase**, **prophase**, **metaphase**, **anaphase**, and **telophase**.



The first stage, **Interphase**, is the period when the cell is not dividing. During this time the cell goes through a cycle of growth, both in size and mass. This phase is called **G1 phase**. Next, the replication of DNA occurs (replication means making a copy). This DNA replication stage is called the **S phase**. There are now twice as many chromosomes. If there were 23 pairs of chromosomes, there are now 46 pairs (the number of chromosomes is now $4n$). After completion of DNA replication the cell enters into another growth phase

Figure 2.4. Mitosis.

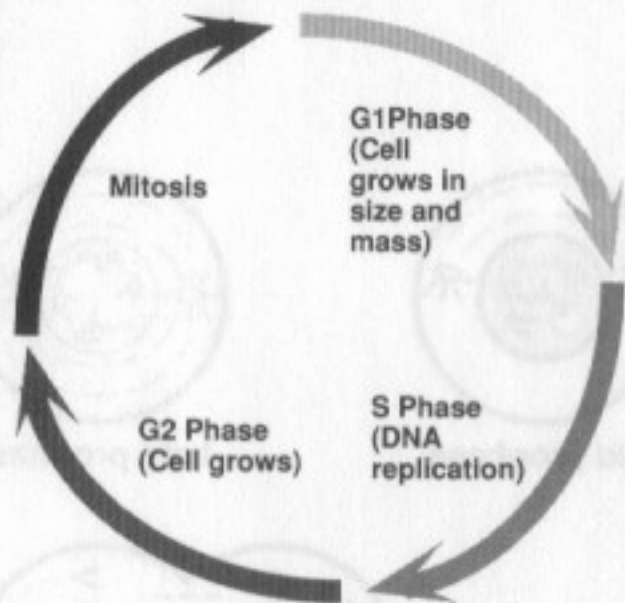


Figure 2.5. Cell cycle.

called **G2 phase**. After G2 phase, the cell is ready to divide.

The cell then divides and enters into the next stage called **prophase**. During this stage the tightly coiled chromosomes start to uncoil. By late prophase, there are two copies of chromosomes which are held together only at narrow constricted areas called **centromeres**. The nuclear membrane breaks down and the two **centrioles** move to the opposite poles of the cell.

The next stage is **metaphase**. During the period of metaphase the chromosomes line up in the middle of the cell.

After metaphase comes **anaphase**. In anaphase the chromosomes separate and move towards the opposite poles of the cell.

The final stage is **telophase**. In telophase, nuclear membranes form around the chromosomes and nucleoli are produced. The cytoplasm of the cell then divides to become two new cells (each cell again contains 23 pairs of chromosomes).

Meiosis

In Mitosis or cell division, each daughter cell inherits one copy of each chromosome pair and maintains the diploid nature of the cell. In contrast, there is another type of cell division called **Meiosis**.

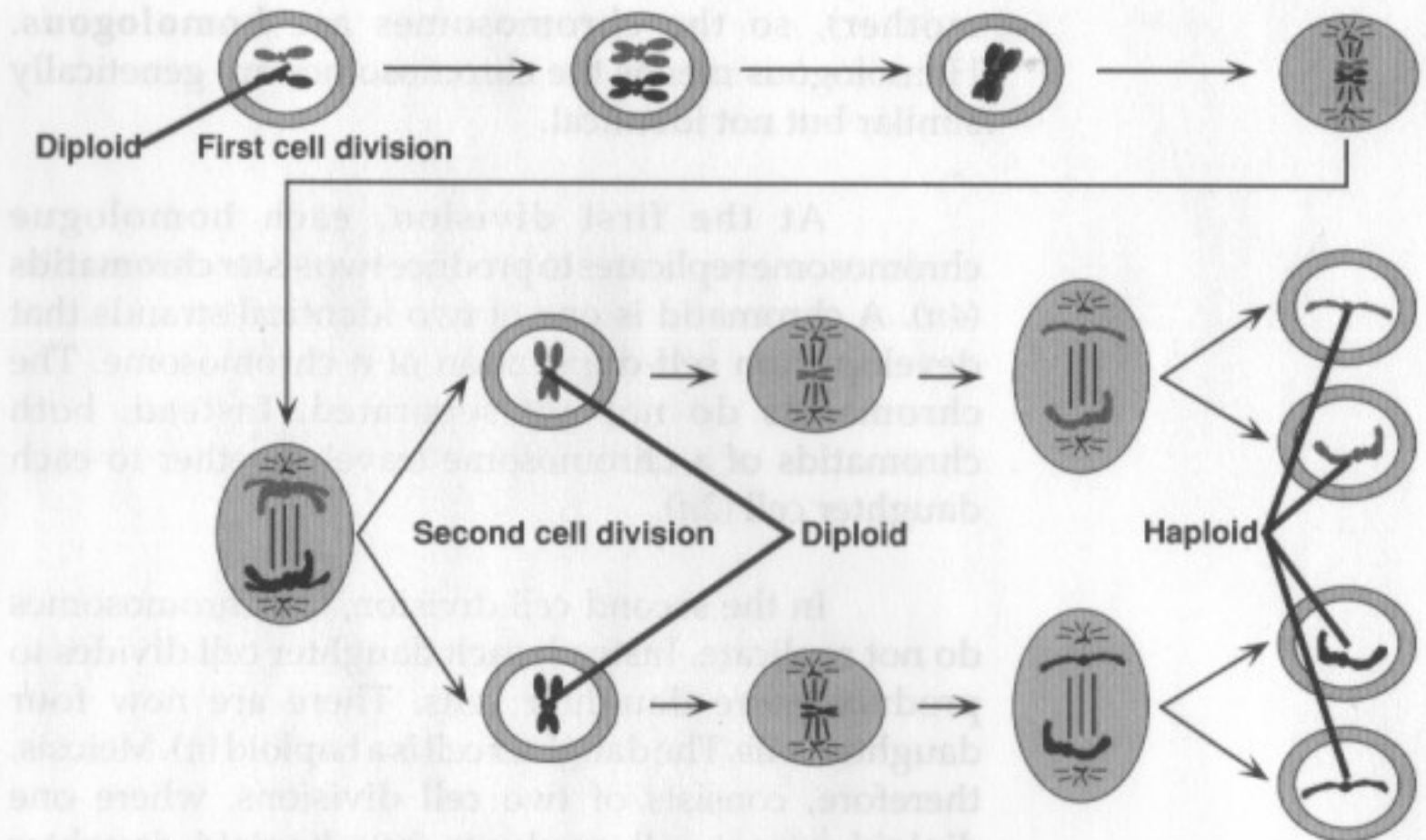


Figure 2.6. Meiosis.

Meiosis occurs when the cell reproduces and divides into gametes. Sperm and ovum are the examples of gametes. Spores of plants are also gametes.

In meiosis, the gamete (daughter cell) inherits only one set of chromosomes (not a pair), either a paternal chromosome (from the father) or a maternal chromosome (from the mother). The gamete (daughter cell) is a **haploid** cell (the number of chromosome is n). It is not a diploid cell.

The cell division in meiosis is similar to mitosis except that it divides twice.

The parent cell is diploid. The two chromosomes are descended from different parents (father and

mother), so the chromosomes are **homologous**. Homologous means the chromosomes are genetically similar but not identical.

At the first division, each homologue chromosome replicates to produce two sister **chromatids** ($4n$). A chromatid is one of two identical strands that develop from self-duplication of a chromosome. The chromatids do not get separated. Instead, both chromatids of a chromosome travel together to each daughter cell ($2n$).

In the second cell division, the chromosomes do not replicate. Instead, each daughter cell divides to produce more daughter cells. There are now four daughter cells. The daughter cell is a haploid (n). Meiosis, therefore, consists of two cell divisions, where one diploid parent cell produces four haploid daughter cells.

Chapter 3

Proteins and Amino acids

Human, animal, and plant cells contain **carbohydrates, fats, proteins and nucleic acids**.

Carbohydrates are substances that are commonly called sugar and starches. Carbohydrates provide energy to the body.

Fats provide heat and insulation to the body. They also provide energy reserves.

Proteins are the most abundant of the organic compounds and play many important roles in the body's function and structure. Consider some of the roles played by different proteins:

1. **Structural proteins:** Structural proteins form support and building blocks for the body. For example, collagen is the protein that provides support for bone and cartilage.

2. **Contractile proteins:** Actin and myosin are proteins present in the muscles. They provide the elasticity necessary to contract the muscles and produce movements.

3. **Enzymes:** Enzymes are proteins that catalyse reactions that store and release energy (to

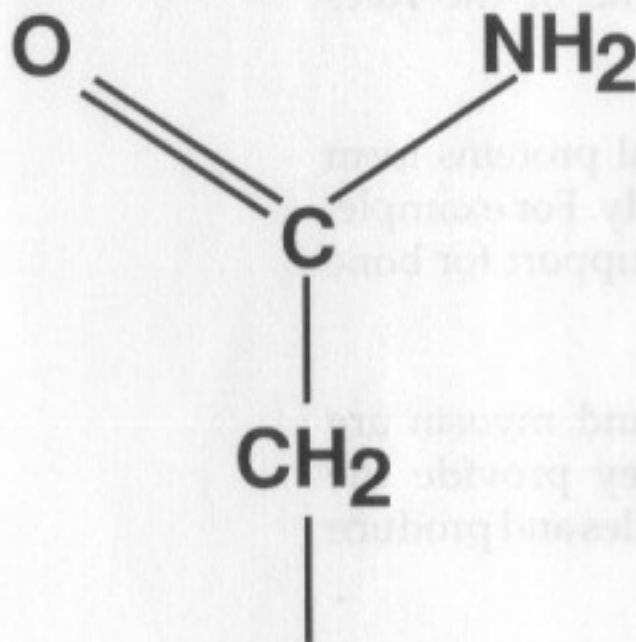
catalyse means to speed up chemical reactions). Enzymes also help in synthesis and breakdown of products in the body.

4. Carrier Proteins: Carrier proteins help in the transportation of molecules. For example, hemoglobin, which is present in the red blood cell, is a protein compound that carries oxygen in the bloodstream to different cells of the body.

5. Antibodies: Antibodies are proteins that protect against infections and injury.

6. Hormones: Hormones are chemical substances that are secreted by the cells and regulate the body metabolism. Metabolism is a process where the food and other substances are built or broken down in the body. Some hormones are proteins or polypeptides.

For example, growth hormone is a polypeptide.



7. Regulatory proteins: Regulatory proteins regulate other genes and the synthesis of other proteins.

The units or building blocks that compose proteins are **amino acids**. Shown (figure 3.1) is a structure of Asparagine, one of the amino acids present in the body.

Figure 3.1 Amino acid Asparagine.
(C=carbon; H=hydrogen;
O=oxygen; N=nitrogen)

When amino acids are linked together by a special chemical bond called **peptide bond** they form a **polypeptide** (figure 3.2).

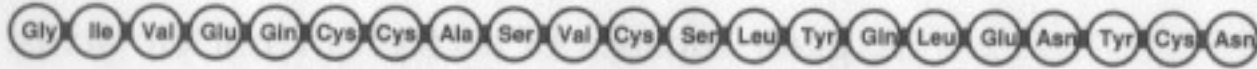


Figure 3.2. Structure of a polypeptide chain.

A protein is made up of one or more polypeptides. For example, insulin, which is a hormone produced by the pancreas and controls the sugar level in the blood, consists of two polypeptides, one of them 21 amino acids long and the other 30 amino acids long (figure 3.3).

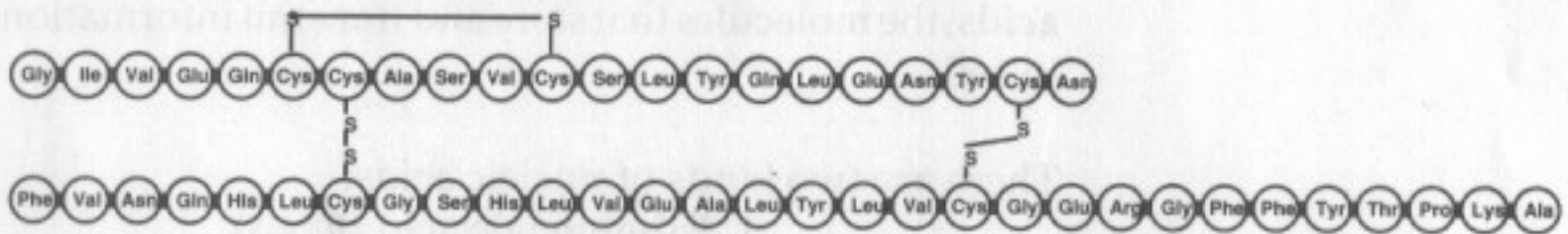


Figure 3.3 Structure of an insulin protein.

The sequence of amino acids is different in each kind of protein molecule. For example, the amino acid sequence present in insulin is different from the amino acid sequence present in thyroid hormone.

There are about 300 amino acids that exist in nature but only 20 of these occur in proteins.

The function of a protein depends upon the sequence of the amino acids. The function and structure of a cell depends upon the types of proteins present. Protein synthesis is, therefore, the most important aspect of cell function. We will discuss about how proteins are synthesized in chapter five.

Nucleic acids: Since the nucleic acids are important components of the body, we will discuss them in the chapter on genes.

Chapter 4 **Genes**

A **gene** consists of a chain of nucleic acids. It contains important biological information which is essential for all living organisms, including plants. There are many genes on each chromosome.

Cells receive instructions about which proteins to synthesize and in what quantities from the **nucleic acids**, the molecules that store and transmit information in cells.

There are two kinds of nucleic acids:

1. **Ribonucleic acid (RNA)**
2. **Deoxyribonucleic acid (DNA).**

Each nucleic acid is a long polymeric molecule consisting of millions of individual units called **nucleotides**. A **polymeric molecule** is a long chain-like molecule with numerous individual units, called **monomers**, linked together. A nucleotide is a monomer. When nucleotides are linked together into a polymeric molecule it is called a **polynucleotide**.

Nucleotides are the basic units of nucleic acids.

Each nucleotide has three components:

1. A **sugar**: The sugar component of the nucleotide has five carbon atoms. It is therefore called a pentose sugar (glucose has six carbon atoms and is called a

hexose sugar). The pentose sugar present in RNA nucleic acid is called ribose, hence the name ribonucleic acid. The pentose sugar in DNA is called deoxyribose and the nucleic acid is therefore called deoxyribonucleic acid (figure 4.1).

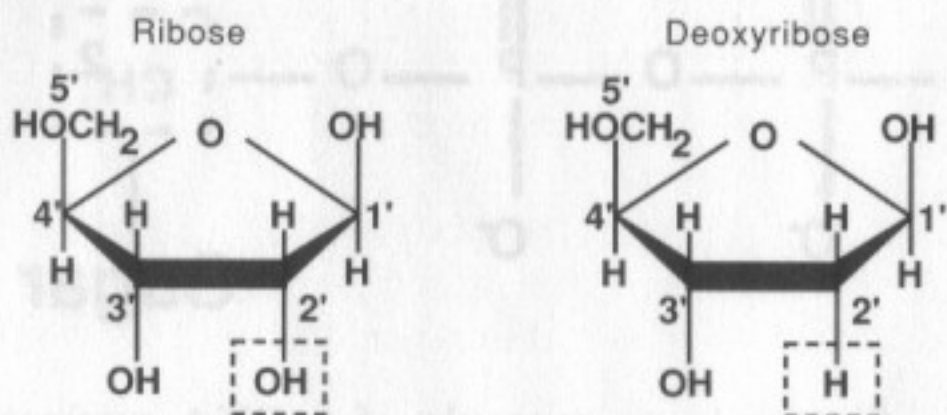


Figure 4.1. Structure of a sugar.
(H=hydrogen; O=oxygen;
C=carbon)

2. A carbon compound containing nitrogen (called nitrogenous base): There are five different nitrogenous bases adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) (figure 4.2). Adenine, guanine, and cytosine can be present in both RNA and DNA. However, thymine is present only in DNA and uracil is found only in RNA.

3. A phosphate group (figure 4.3).

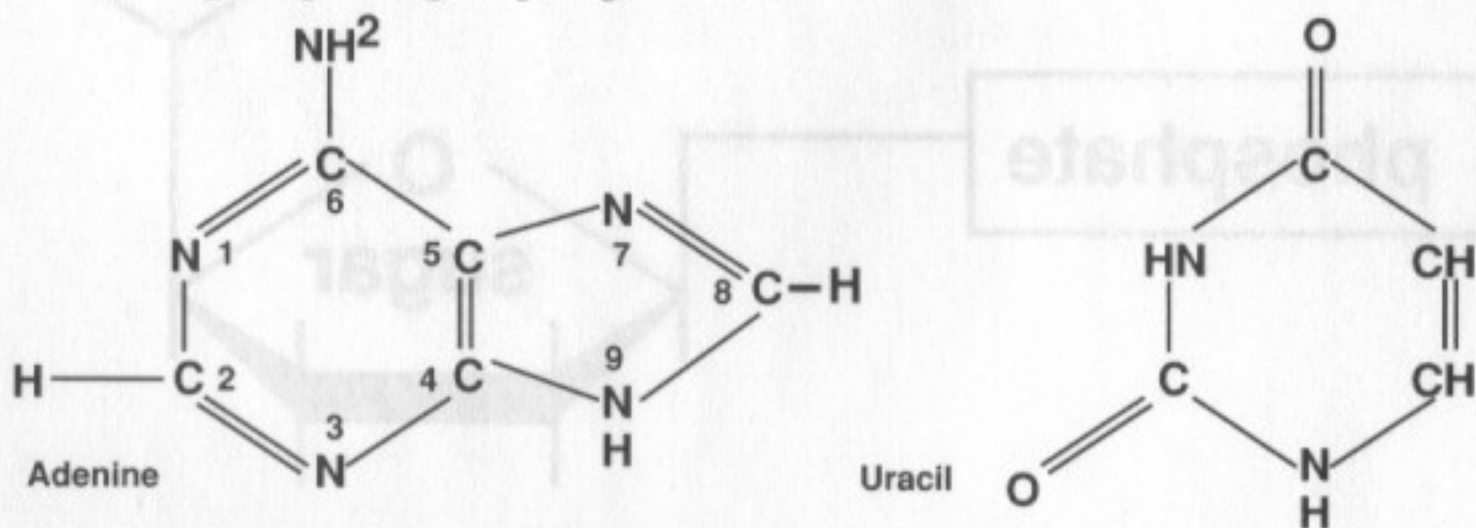


Figure 4.2. Structure of a nitrogenous base. (H=hydrogen; O=oxygen; C=carbon; N=nitrogen)

A nucleotide has sugar, base, and phosphate linked together (figure 4.4).

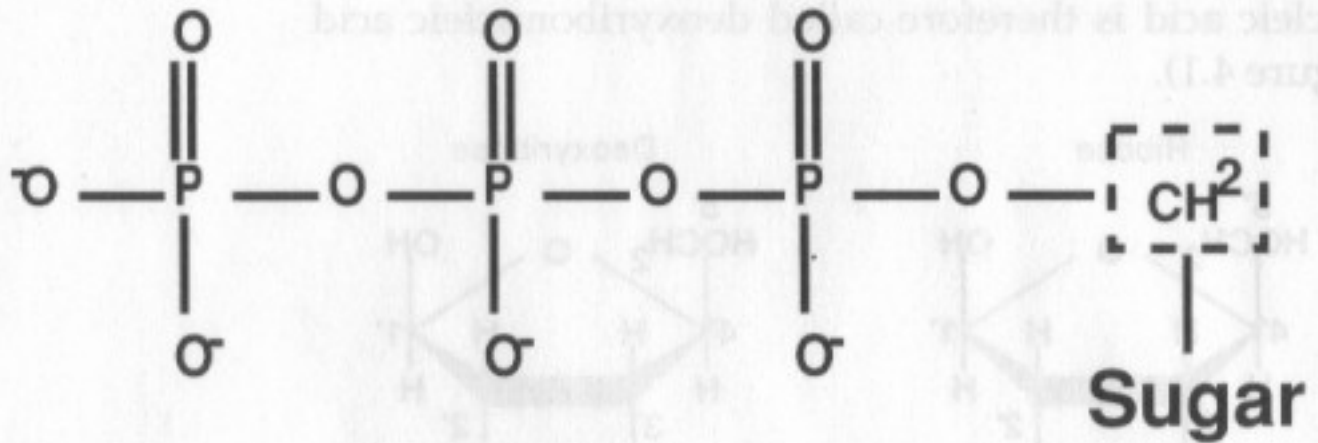


Figure 4.3. Structure of a phosphate.

Shown is an example of a DNA nucleotide containing the nitrogenous base adenine (figure 4.5).

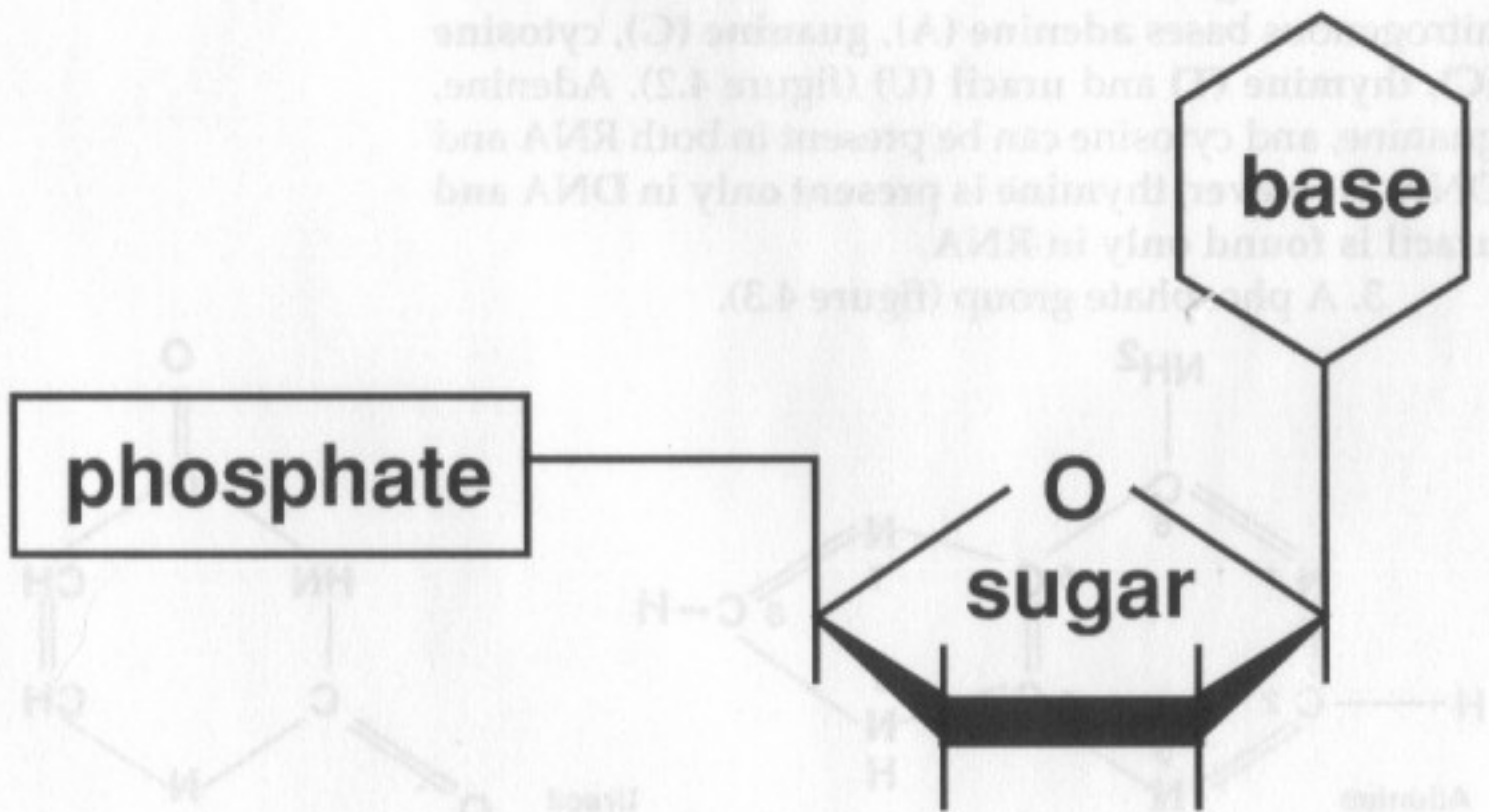
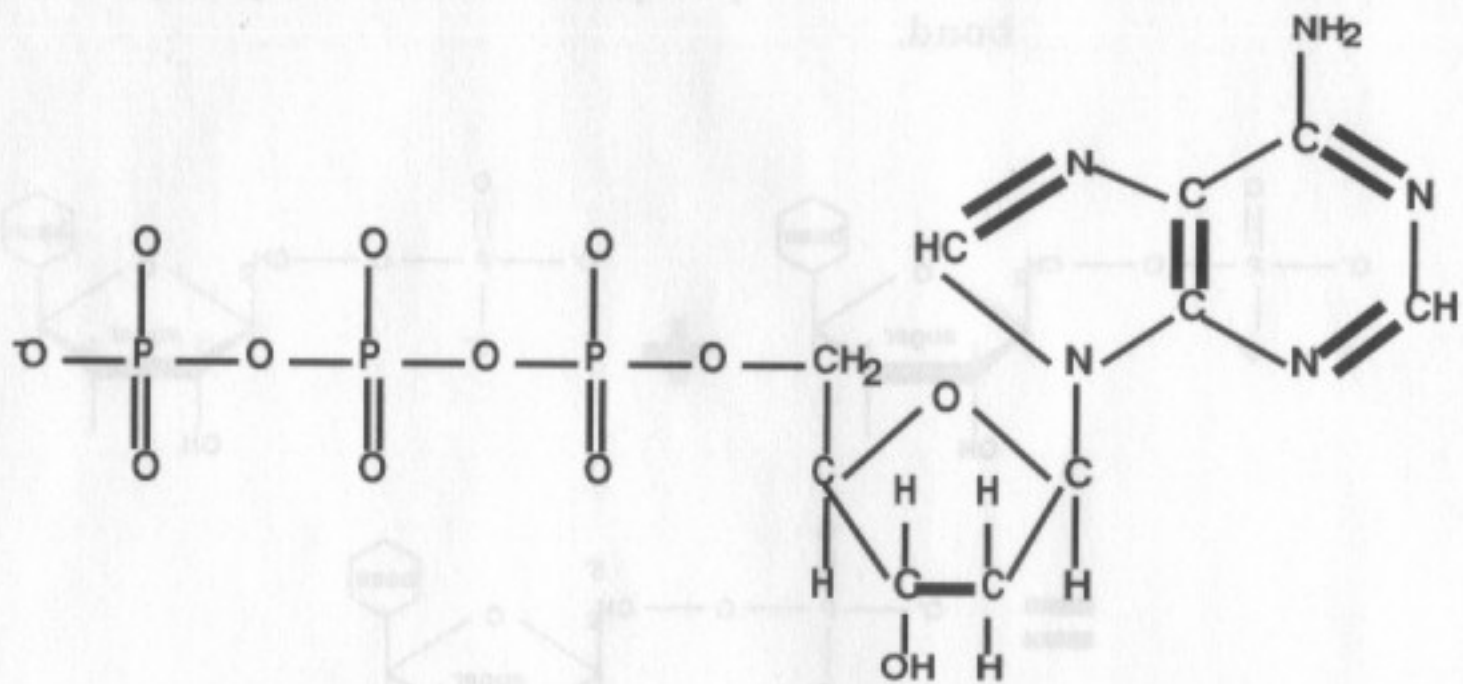


Figure 4.4. Structure of a nucleotide.

The nucleotide is called 2'-deoxyadenosine 5'-triphosphate (in short form it is just labeled A).

The other nucleotides are:

2'-deoxyguanosine 5'-triphosphate (G)



2'- deoxycytidine 5'-triphosphate (C)

2'-deoxythymidine 5'-triphosphate (T).

(Note: the numbers that are seen on the base and the sugar. These numbers on the nitrogenous bases are labeled 1,2,3,4,5. The numbers on the sugar are labeled 1',2',3',4',5'. The numbers are read as 'one-prime', 'two-prime', 'three-prime', etc.).

Figure 4.5. Structure of a deoxyadenosine nucleotide.

When more than one nucleotide is linked together it becomes a nucleic acid, either a DNA or a RNA depending upon whether the sugar present is deoxyribose (DNA) or ribose (RNA). The following is an example of a small DNA molecule with two

G-NETIX

nucleotides linked together (figure 4.6).

The nucleotides are joined at the phosphates attached to the 5'-carbon of one nucleotide and to the 3'-carbon of the next nucleotide in the chain. The linkage between the phosphates is called a **phosphodiester bond**.

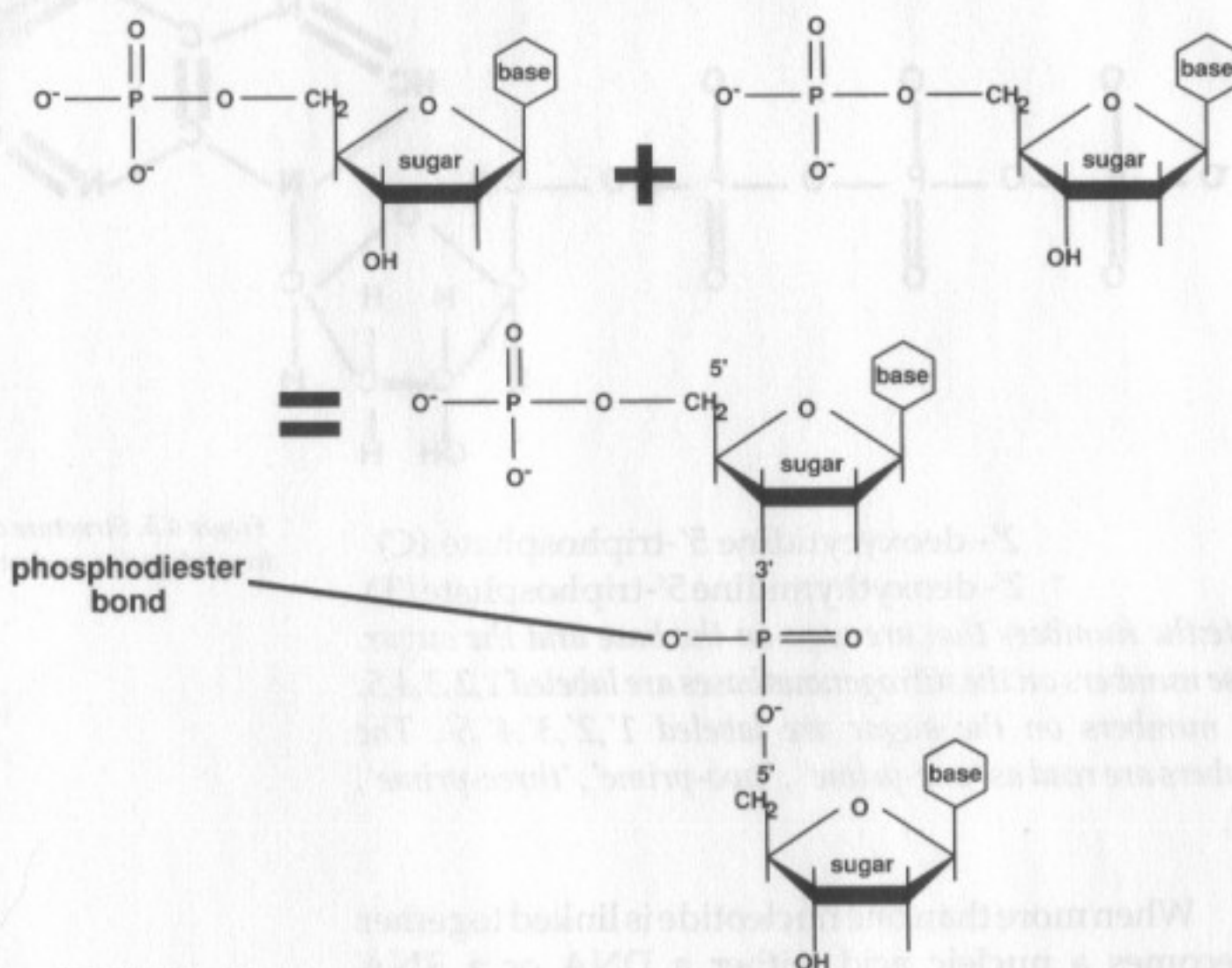


Figure 4.6. Nucleic acid.

The two ends of the DNA molecule are not the same. The top of the molecule ends with a phosphate group attached to the 5'-carbon of the sugar. This end is called 5' or 5' -P terminus. The other end of the DNA

molecule has an -OH attached to the 3'-carbon of the sugar. This is called 3' or 3'-OH terminus.

These are important terms because the sequence of nucleotides in a DNA molecule can be read in 3' to 5' direction (3' -> 5') or 5' to 3' direction (5' -> 3'). Depending upon the direction, the interpretation of the sequence may be different.

The DNA nucleic acid molecule (polynucleotide) may have a few or thousands of nucleotides linked to each other.

Using a special technique called x-ray diffraction analysis, the DNA molecule has been found to be a **helix** (helix means something spiral).

After complicated analysis, two scientists named James Watson and Francis Crick were able to deduce that the DNA exists as **double helix** (you might want to read Watson's fascinating book *Double Helix* to see how they came to this conclusion). In a double helix there are two chains of polynucleotides which spiral around each other. One polynucleotide runs in the 5'-3' direction and the other in the 3'-5' direction. The two polynucleotides in a double helix are therefore **complementary**, the sequence of one nucleotide determines the sequence of the other nucleotide. The polynucleotides from the two chains are connected through the nitrogenous bases by hydrogen bonds. **The connections between the nitrogenous bases are always Guanine and Cytosine (G-C) or Thymine and Adenine (T-A).** (figure 4.7). Each connection between these two nitrogenous bases is called a **base pair** (abbreviated as



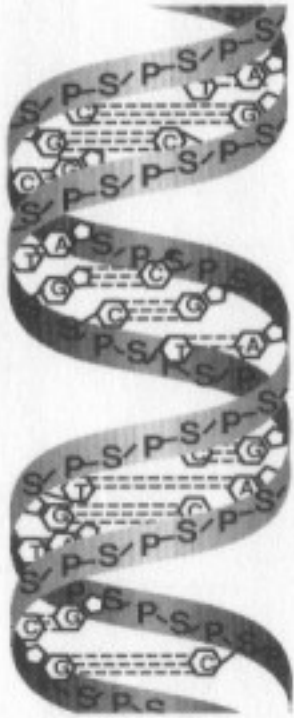


Figure 4.7. DNA double helix.

bp). The length of a DNA is not measured in inches or centimeters. It is measured by the number of base pairs e.g. 10 bp, 100 bp etc.

In addition to the DNA, there is another nucleic acid called **Ribonucleic acid (RNA)**.

RNA differs from DNA in that:

1. The sugar in RNA is **ribose** instead of deoxyribose.

2. One of the nitrogenous bases is **uracil** instead of thymine. RNA contains nitrogen bases adenine, guanine, cytosine and uracil (DNA contains adenine, guanine, cytosine and thymine).

Note: for the more biologically oriented: Adenine and guanine are purines. Cytosine, thymine and uracil are pyrimidines.

3. RNA does not exist as double helix. It exists only as a single polypeptide chain.

The nucleotides present in RNA are:

- adenosine 5'-triphosphate (A)
- guanosine 5'-triphosphate (G)
- cytidine 5'-triphosphate (C)
- uridine 5'-triphosphate (U).

There are three types of RNA:

- messenger RNA (mRNA)**
- transfer RNA (tRNA)**
- ribosomal RNA (rRNA)**

Messenger RNAs (mRNA) act as intermediaries between the genes and the polypeptide products. They take the information from the genes (a process called **transcription**) and act as templates for the synthesis of the polypeptide chains (a process called **translation**). mRNAs usually last for a few minutes in the cell before they are degraded.

Transfer RNAs (tRNA) are small molecules containing about 70 to 90 nucleotides. tRNAs act as adapter molecules that read the nucleotide sequence of the mRNA and convert them into a sequence of amino acids. The shape of a tRNA molecule looks like a cloverleaf (figure 4.8).

Each tRNA has an **anticodon arm** which interprets the sequence of the nucleotides and an **acceptor arm** which attaches to the corresponding amino acid. The anticodon arm has a triplet nucleotide sequence (triplet means three) which is complementary to the codon on the mRNA.

Anticodon is a term used for three nucleotides which are complementary to the codon and **codon** is a term used for the three sequential nucleotides that code for a single amino acid. The amino acid is attached to the acceptor arm of the tRNA (figure 4.9, pg. 28).

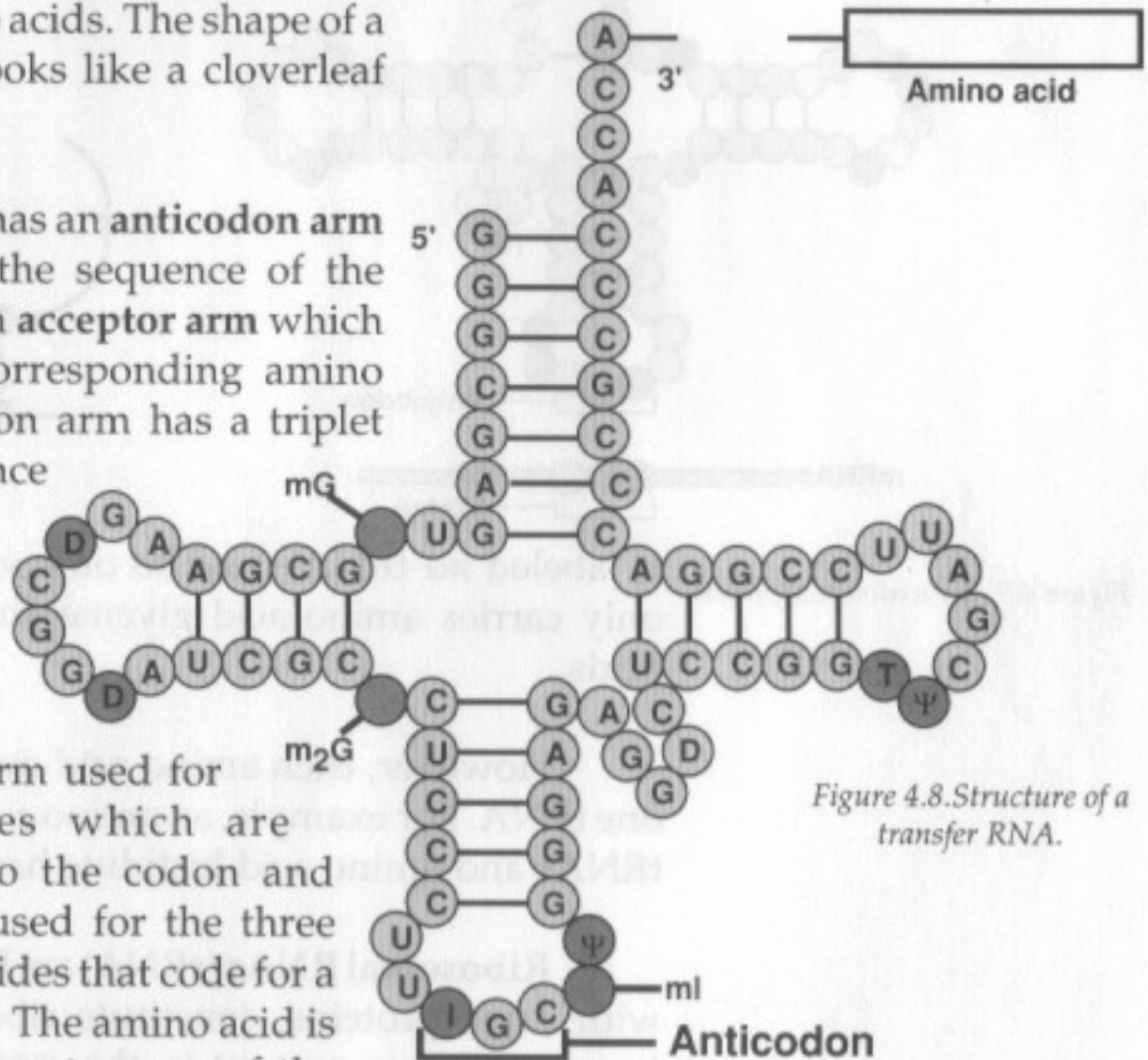


Figure 4.8. Structure of a transfer RNA.

G-NETIX

Each tRNA is designed to carry only one of the 20 amino acids used for protein synthesis. For example, a tRNA that carries an amino acid glycine is labeled as tRNA^{Gly} and a tRNA that carries an amino acid leucine

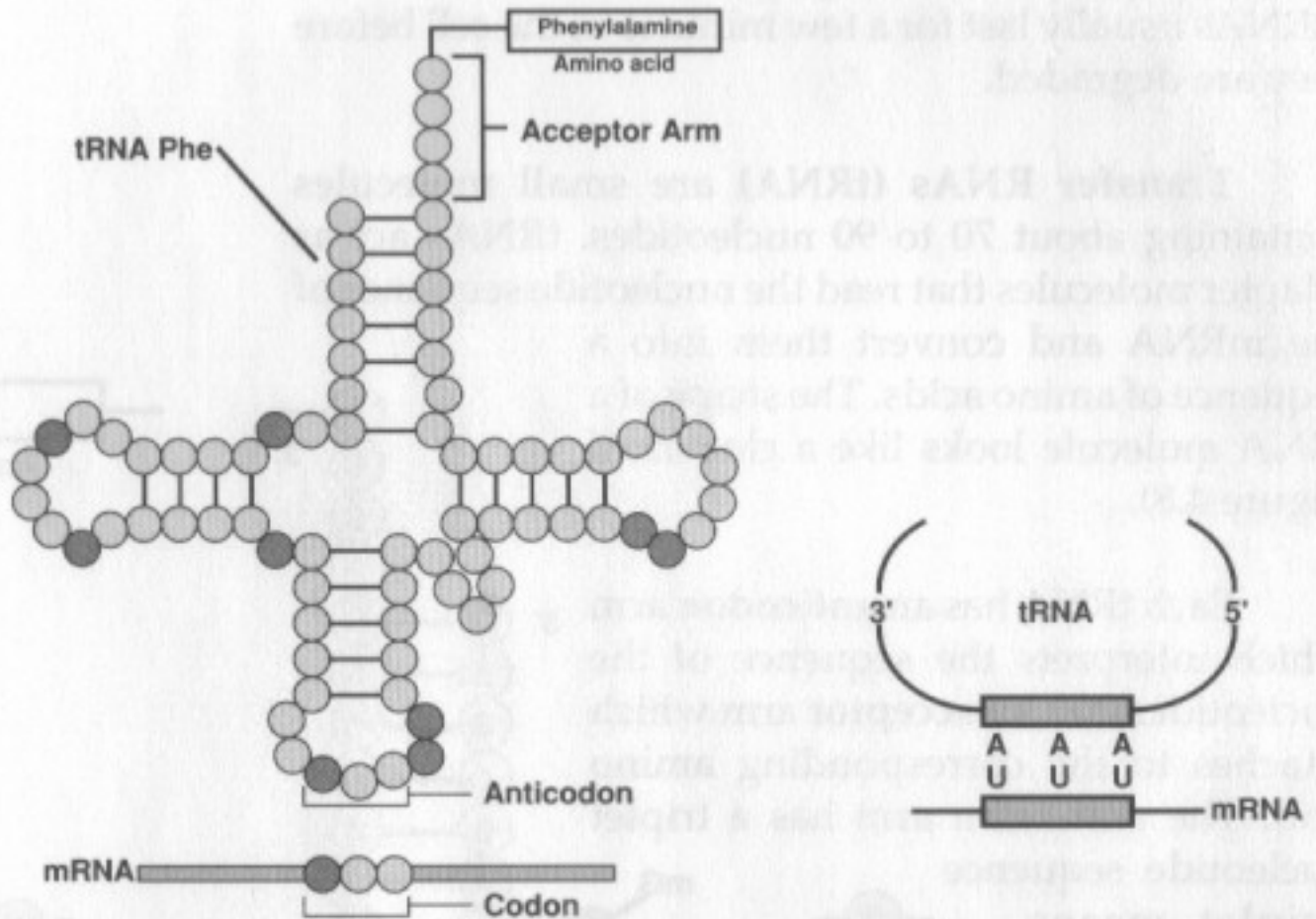


Figure 4.9. Anticodon recognition.

is labeled as tRNA^{leu} and so on and so forth. tRNA^{Gly} only carries amino acid glycine and no other amino acids.

However, each amino acid may have more than one tRNA. For example, an amino acid alanine has four tRNAs and amino acid histidine has two tRNAs.

Ribosomal RNAs (rRNA) are RNAs that, together with some proteins, constitute ribosomes which are large structures present in the cytoplasm of the cell.

Ribosomes are considered as factories that synthesize proteins.

Ribosomes attach to mRNA molecules and move along them, synthesizing polypeptides as they go. (figure 4.10).

How is the information stored?

Genes are made of polynucleotides. A gene is just a segment of the DNA molecule.

The segment may be any length from about 75 nucleotides to hundreds of nucleotides. It is now known that the biological information is contained in the nucleotide sequence of the gene. This biological information is carried by just one of the two polypeptides of the double helix.

The strand of polypeptides that contains the gene information is called the **template** or **non-coding strand**. The other strand is called **coding** or **complementary strand**. At various times the template strand for one gene may also become a complementary strand for another gene (figure 4.11, pg. 30). The sequence of nucleotides in a gene in the template strand is read in the 3' to 5' direction. If the sequence is read in the opposite direction i.e. 5' to 3' direction, the interpretation will be wrong.

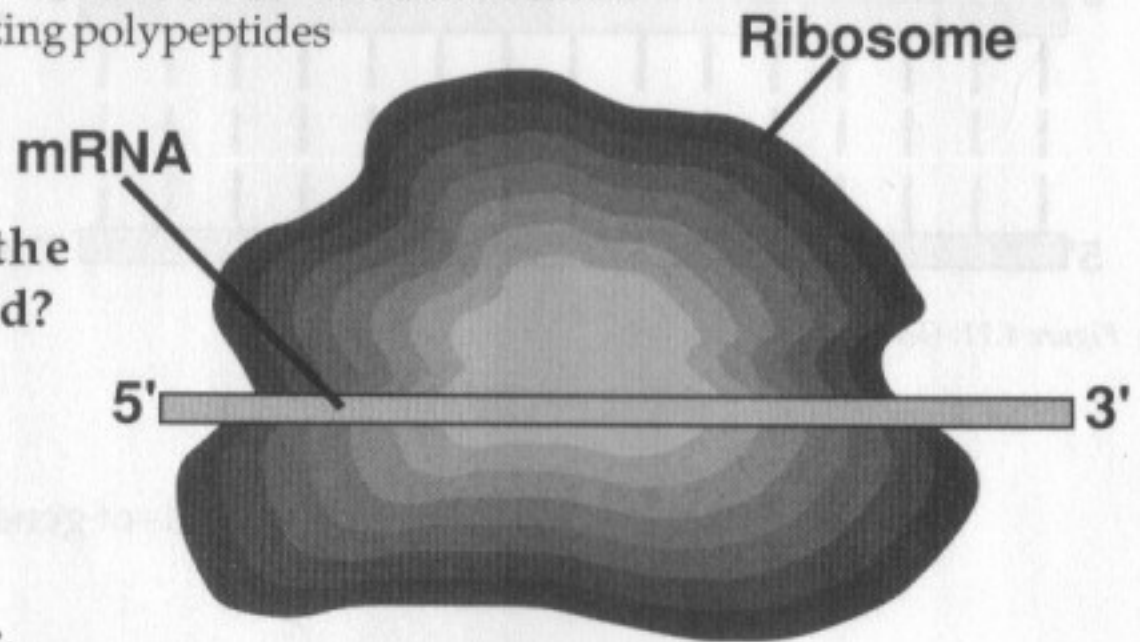


Figure 4.10. Ribosomal RNA.

Some of the genes may occur in groups (clusters). These genes contain related information.

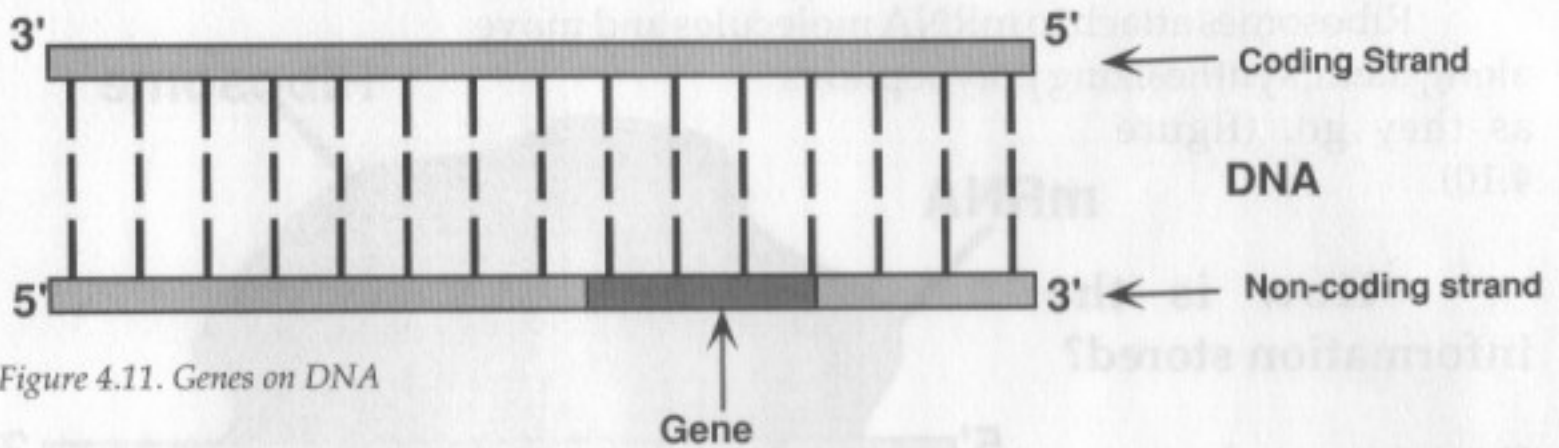


Figure 4.11. Genes on DNA

There are two kinds of gene clusters, **operon** and **multigenes**.

An **operon** is a group of genes that code for a series of enzymes. All these enzymes must be present to complete a biochemical process. Therefore, all the genes in the operon must be operational for a reaction to occur. Operons are not seen in any other organisms except bacteria (that does not mean it is not present in the game).

In a bacteria called *E. coli*, there is a cluster of three genes that encode three enzymes respectively. These three enzymes work together to convert a sugar called lactose into other sugars called glucose and galactose (figure 4.12, pg. 31). The cluster of genes that code the enzymes is an operon. These three genes that encode the enzymes normally do not become active. However, the operon genes become active when there is lactose present in the environment.

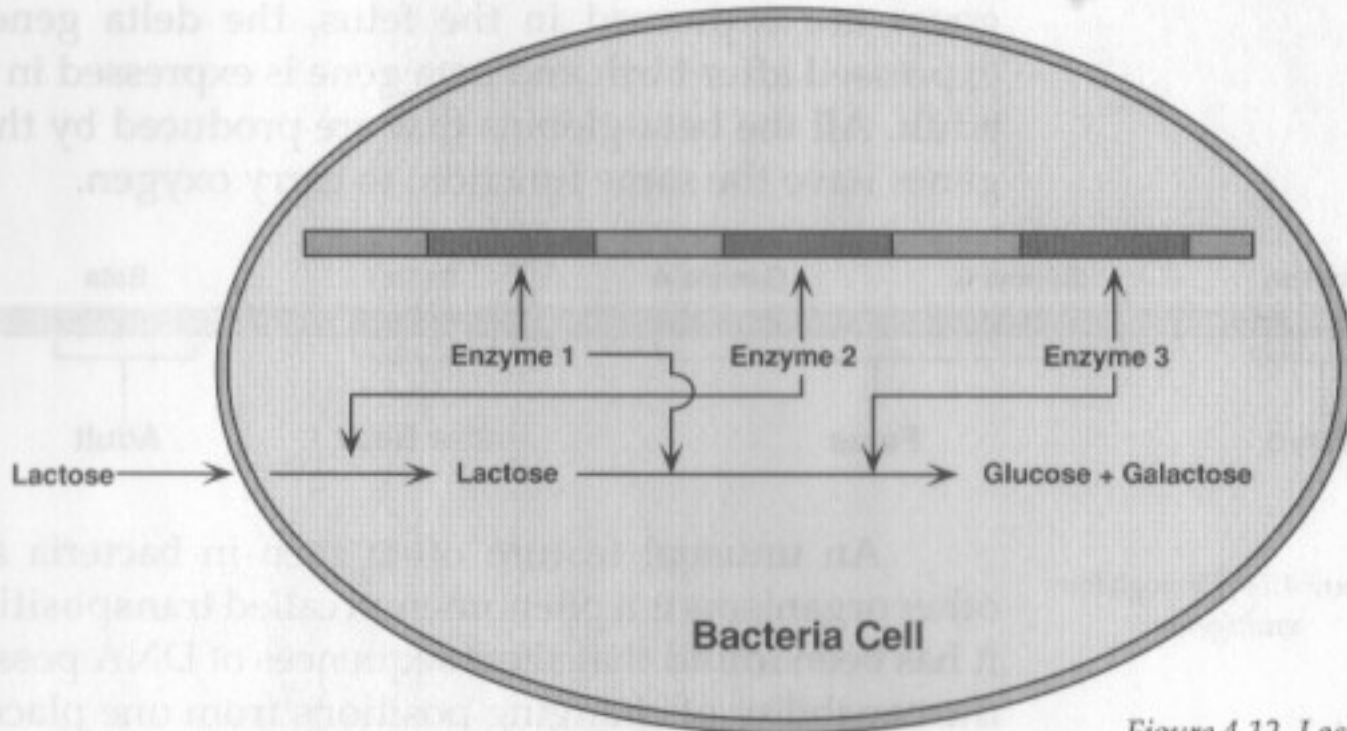


Figure 4.12. Lactose operon.

Multigene families are a cluster of genes that are related. The nucleotide sequences of the individual genes in the multigene family are nearly identical. The function of these genes are also similar.

In humans and animals there is an iron-protein complex called hemoglobin. Hemoglobin carries oxygen from the lungs to other organs and tissues in the body. Hemoglobin is an example of multigenes. Hemoglobin contains two pairs of globin proteins (or polypeptides), alpha and beta.

There are five beta-globin genes differing from one another by a few amino acid gene sequence only (figure 4.13, pg. 30). They are beta, delta, gamma A, gamma G, and epsilon genes. The genes are located on the chromosome 11. Each gene expresses itself at different time during the development of an embryo.

The epsilon gene is expressed in the embryo, the gamma genes are expressed in the fetus, the delta gene is expressed after birth and beta gene is expressed in the adult. All the beta-globins that are produced by these genes have the same function, to carry oxygen.

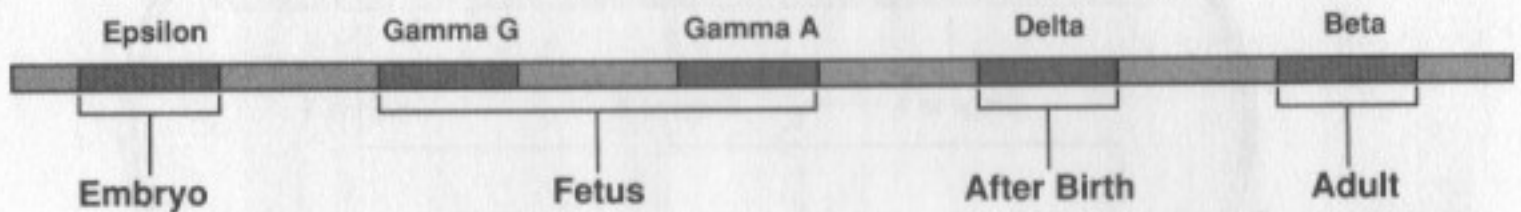


Figure 4.13. Hemoglobin multigene.

An unusual feature often seen in bacteria and other organisms is a phenomenon called **transposition**. It has been found that short sequences of DNA possess the capability of changing positions from one place to another. These short sequences are called **transposable elements** or simply **transposons**. When the transposon changes its position, it can result in a mutation of the gene it has departed from as well as the gene it has inserted into (figure 4.14).

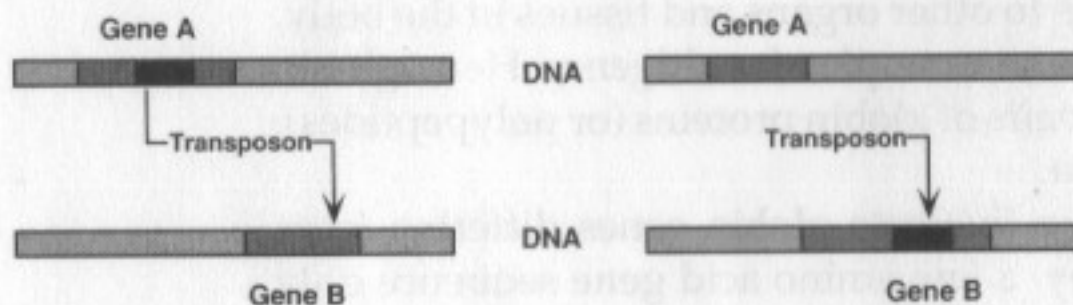


Figure 4.14. Transposons.

Chapter 5

How proteins are formed

From the previous chapter we know that proteins are the most important building blocks of all living organisms. The characteristics of each species is determined by the composition of its proteins. Different proteins interact with each other to form complex organisms such as humans, animals, and plants.

We also know that proteins are composed of different amino acids linked together in a chain. The way that amino acids are linked to form proteins is regulated by the genes. The nucleotide sequence of a gene determines the amino acid sequence of a protein.

This synthesis of protein occurs in two stages (fig. 5.1, pg 34):

1. **Transcription:** During transcription the information stored in the gene is transferred to an RNA molecule (called **messenger RNA** or **mRNA**) which carries the information from the genes to the **ribosomes**, the sites for protein synthesis.

2. **Translation:** During translation, mRNA acts as a template for the synthesis of a polypeptide and, therefore, a protein. Translation occurs in the ribosomes.

Transcription

In transcription, one of the strands (the **non-coding strand**) of the DNA double helix becomes a template for the synthesis of mRNA.

G-NETIX

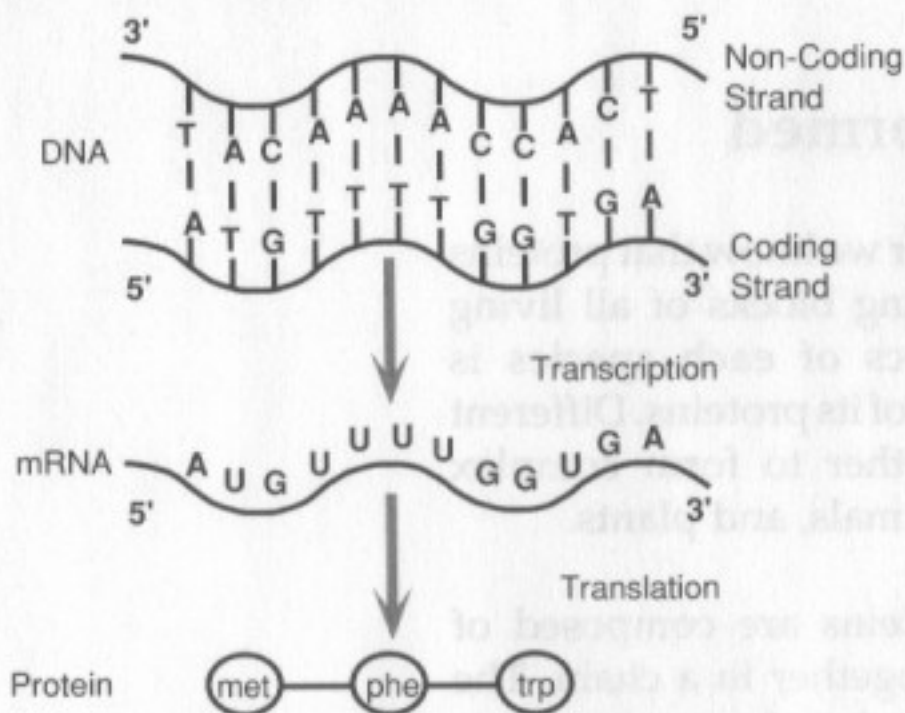


Figure 5.1. Transcription and translation.

Not all of the nucleotide sequence in the DNA gives biological information. Some genes are interrupted by nonsense sequences called **introns**. The sequence of nucleotides that does provide information is called an **exon** (figure 5.2). An exon is also called a **coding sequence** since it provides necessary information for coding.

During transcription an enzyme called RNA polymerase binds to the DNA at a specific position called a **promoter**. In eukaryotes, different RNA polymerase molecules synthesize different types of RNA.

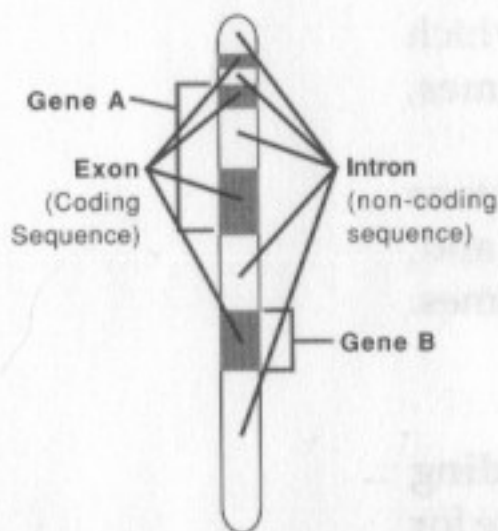


Figure 5.2 Introns and exons

A promoter is a short nucleotide sequence that is recognized by an RNA polymerase enzyme as a beginning point (start site) for mRNA synthesis. All promoters have the same or similar nucleotide sequences if each is to be recognized by the same enzyme.

A **consensus sequence** is a sequence of nucleotides that are similar in many different genes. From the information obtained from a bacterium called *E. coli*, the consensus sequence of the promoter is found at two locations. The **-10 sequence** (also known as the **Pribnow box** after

the scientist who discovered it) is located ten nucleotides before the starting point of transcription and is most commonly **TATAAT**. The **-35 sequence** is located thirty-five nucleotides before the starting point of transcription and is usually **TTGACA**.

In eukaryotes, the promoters for another kind of polymerase called RNA polymerase II have different sequences at different locations e.g. **GGNNCAATCT** (called the **CAAT box**) at the **-75** location and **TATAAAT** (called the **Hogness box** or the **TATA box**) at the **-25** location).

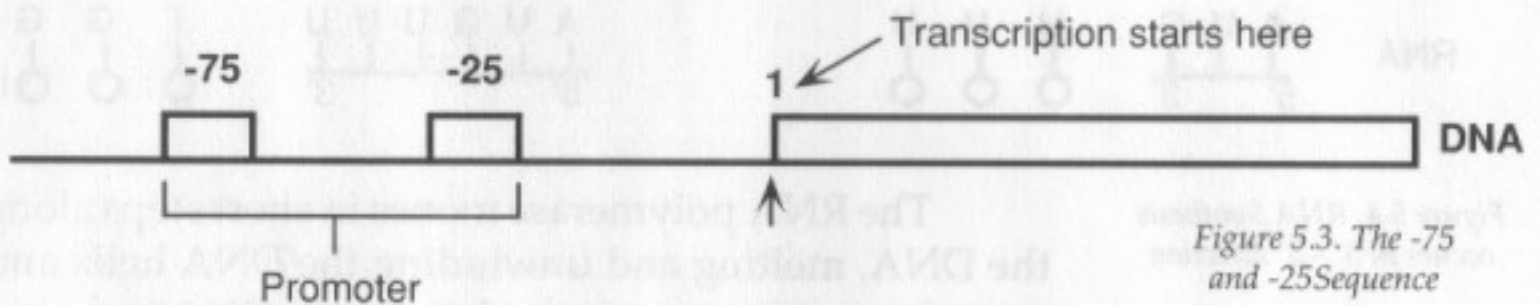


Figure 5.3. The **-75** and **-25** Sequence

After binding to the promoter, the RNA polymerase unwinds the DNA strands and exposes the nucleotides which are present at that site.

One of the two exposed strands, the **non-coding strand**, acts as a template where the nucleotides of the RNA attach (remember the nucleotide here is the RNA nucleotide not the DNA nucleotide).

The attachment of RNA nucleotides to DNA nucleotides is complementary. For example, if the nucleotide on the DNA is guanosine, the complementary nucleotide on the newly synthesized RNA is cytosine. If

G-NETIX

the nucleotide on the DNA is adenine, the nucleotide on the RNA is uracil (Not thymidine: thymidine is present only in the DNA). The synthesis of RNA always occurs in the 5' to 3' direction. Each new nucleotide is attached to the 3' end of the growing RNA molecule. Because the synthesis of RNA occurs in 5' to 3' direction, the DNA template is read in the 3' to 5' direction by the RNA polymerase.

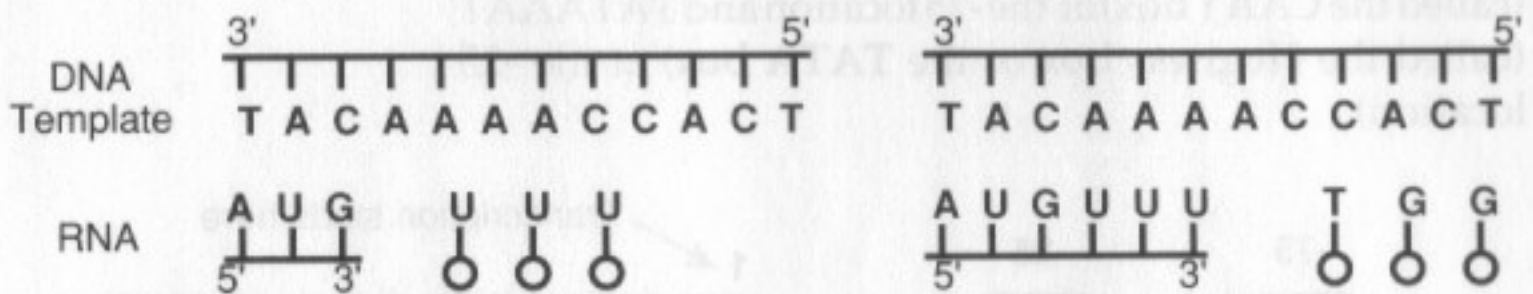


Figure 5.4. RNA Synthesis occurs in 5' - 3' direction

The RNA polymerase moves in short steps along the DNA, melting and unwinding the DNA helix and exposing a new area of template for the RNA to form a complementary pairing.

The melting and unwinding process continues until the RNA polymerase encounters a special sequence in the DNA called a **termination sequence**. At this point, the polymerase releases both the DNA template and the newly made RNA chain. The **termination sequences possess a common feature: they are complementary palindromes** (A palindrome is a sequence that looks the same whether it is read backward or forward. For example, GGTCTTGG is a palindrome).

*Note: some transcription-termination signals require the presence of a protein called **rho**, whereas others do not.*

Each completed RNA chain is released from the DNA template as a free, single stranded RNA molecule.

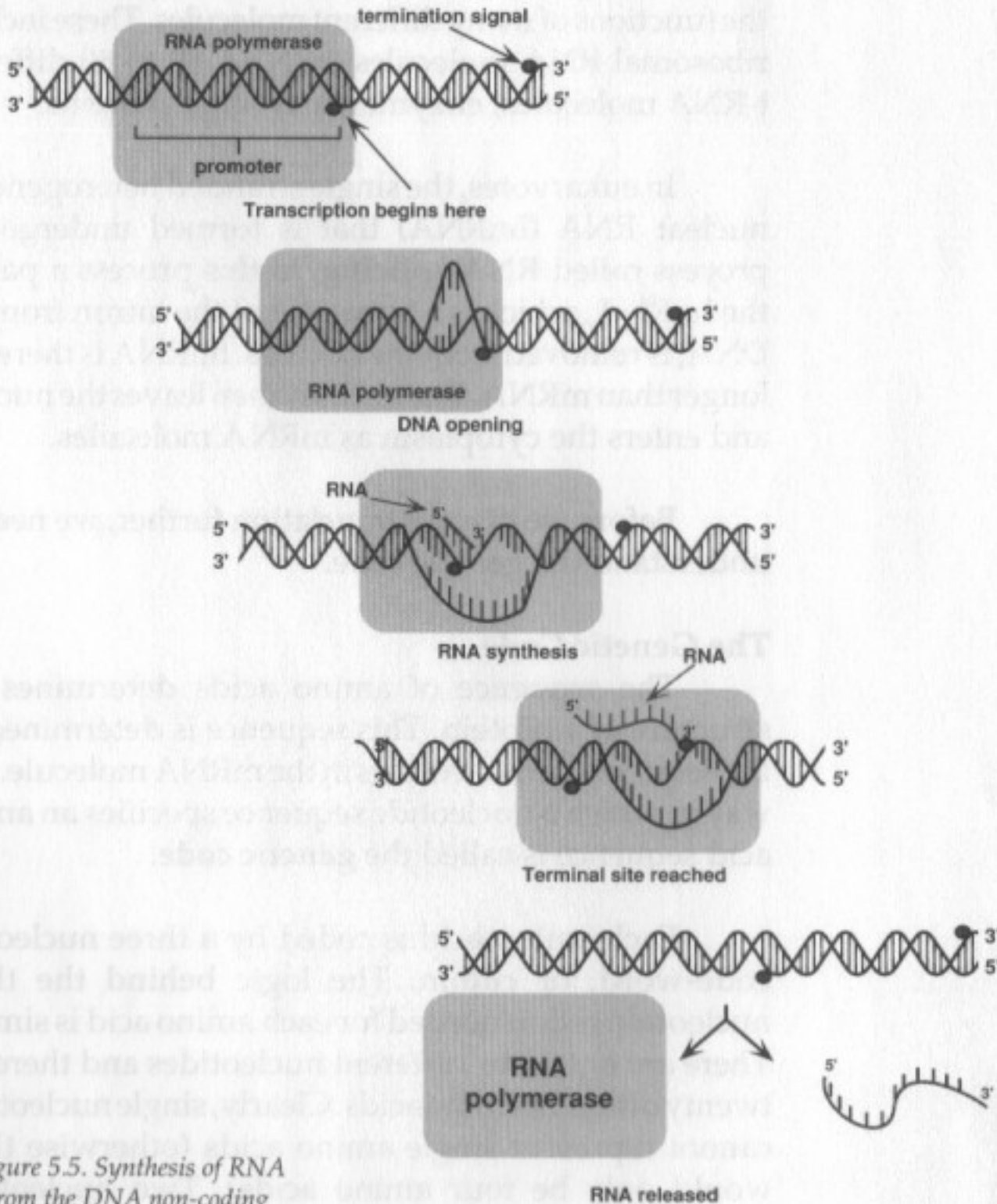


Figure 5.5. Synthesis of RNA from the DNA non-coding strand.

This RNA, called **heterogeneous nuclear RNA (hnRNA)**, later becomes mRNA.

Once transcription is completed, translation occurs. The process of translation is complex, requiring the functions of many different molecules. These include ribosomal RNA molecules (rRNA), 40 to 60 different t-RNA molecules, enzymes, and other proteins.

In eukaryotes, the single stranded heterogeneous nuclear RNA (hnRNA) that is formed undergoes a process called **RNA splicing**. In this process a part of the hnRNA, which is a transcript of the intron from the DNA, is removed from the nucleus. hnRNA is therefore longer than mRNA. The hnRNA then leaves the nucleus and enters the cytoplasm as mRNA molecules.

Before we discuss translation further, we need to understand the genetic code.

The Genetic Code

The sequence of amino acids determines the structure of a protein. This sequence is determined by the sequence of nucleotides in the mRNA molecule. The way in which a nucleotide sequence specifies an amino acid sequence is called the **genetic code**.

Each amino acid is coded by a three nucleotide code-word, or **codon**. The logic behind the three nucleotide codon needed for each amino acid is simple. There are only four different nucleotides and there are twenty different amino acids. Clearly, single nucleotides cannot represent single amino acids (otherwise there would only be four amino acids). Two nucleotides

(such as AT, TA, CG, GG, etc.) do not work either. This would only represent sixteen amino acids ($4^2 = 16$). However, three nucleotides (ATG, GGG, CGA, etc.) work, as this allows for sixty-four code-words ($4^3 = 64$).

Some amino acids are coded for by more than one codon. When an amino acid is coded by more than one codon the genetic code is said to be **degenerate**.

The following table shows the amino acids with their respective mRNA codons.

Alanine	Ala	GCA	GCC	GCG	GCU		
Arginine	Arg	AGA	AGG	CGA	CGC	CGG	CGU
Asparagine	Asn	AAC	AAU				
Aspartic Acid	Asp	GAC	GAU				
Cysteine	Cys	UGC	UGU				
Glutamic Acid	Glu	GAA	GAG				
Glutamine	Gln	CAA	CAG				
Glycine	Gly	GGA	GGC	GGG	GGU		
Histidine	His	CAC	CAU				
Isoleucine	Ile	AUA	AUC	AUU			
Leucine	Leu	UUA	UUG	CUA	CUC	CUG	CUU
Lysine	Lys	AAA	AAG				
Methionine	Met	AUG					
Phenylalanine	Phe	UUC	UUU				
Proline	Pro	CCA	CCC	CCG	CCU		
Serine	Ser	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	ACA	ACC	ACG	ACU		
Tryptophan	Trp	UGG					
Tyrosine	Tyr	UAC	UAU				
Valine	Val	GUA	GUC	GUG	GUU		

Three of the sixty-four codons (**UAA, UAG, UGA**) do not code for any amino acid. Instead they signal the termination of a polypeptide chain. They are known as **stop codons**. When one of the stop codons is reached, the synthesis of the polypeptide stops. Therefore, there are sixty-one codons that code for the twenty amino acids and three stop codons for a total of sixty-four codons.

On the other hand, the codon **AUG** virtually always occurs at the start of a gene and marks the position where translation should begin. **AUG** is therefore called the **initiation codon**. **AUG** is also a **code for the amino acid methionine**. Most of the newly synthesized polypeptides will have this amino acid at the beginning although it may be removed later.

Figure 5.6 is an example of a polypeptide chain with the corresponding mRNA codes.

Polypeptide	Pro - Val - Leu - His - Trp - His
mRNA	CCC-GUC-CUC-CAU-UGG-CAC

Figure 5.6.

Translation

In eukaryotes, the translation process occurs on **ribosomes**. These are complex structures of rRNA and proteins that are present in the cytoplasm.

A ribosome contains **three binding sites: one for mRNA and two for tRNAs**. One of the sites for tRNAs is called the **peptidyl-tRNA binding site**, or **P-site**. A P-site binds the tRNA which is attached to the growing polypeptide chain.

Another site is called the **aminoacyl-tRNA binding site**, or **A-site**. An A-site binds the tRNA carrying the amino acid that is to be added to the polypeptide chain. A small subunit of the ribosome attaches to the mRNA.

(Note: The ribosomal attachment to the mRNA occurs at a specific site just upstream of the initiation codon of the gene. In certain bacteria it has a consensus sequence 5'-AGGAGGU-3' and is called *Shine-Dalgarno sequence*).

Once the ribosome attaches to the mRNA at the binding site it migrates along mRNA until it encounters an initiation codon which consists of the nucleotides AUG. This is where a special tRNA, the **initiator tRNA** (also known as tRNA^{met}) attaches. Remember that the codon AUG represents the amino acid methionine. Therefore, all polypeptides begin with the amino acid methionine. However, methionine is subsequently removed from the polypeptides. Thus, the final polypeptide may or may not have methionine as a beginning amino acid.

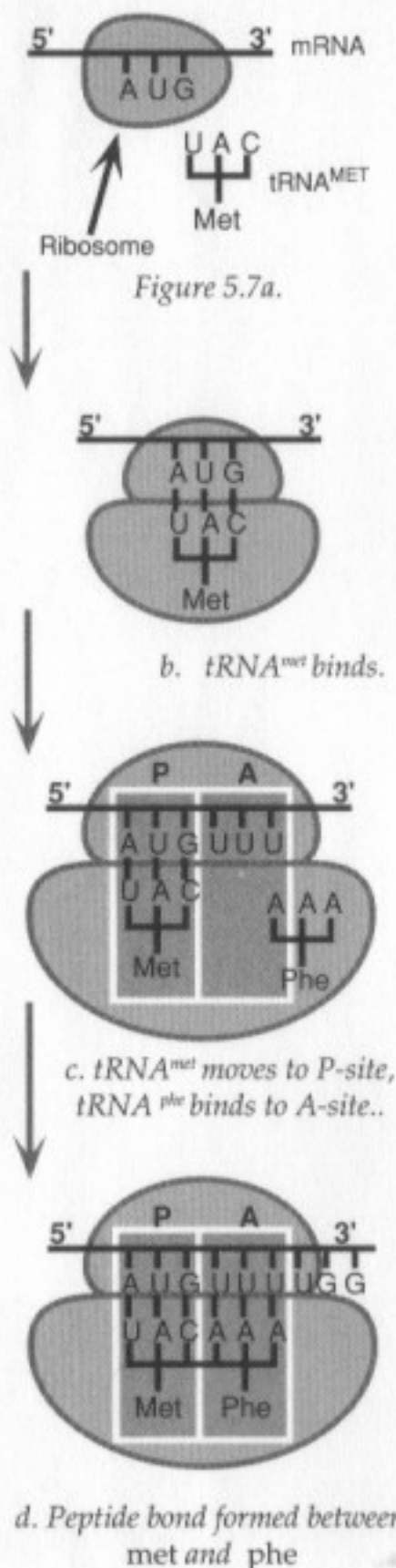
How the initiator tRNA^{met} is distinguished from tRNA^{met} in the middle of a gene is not well known. Discussion of how the translation process differentiates AUG as an initiation codon or as the amino acid methionine is beyond the scope of this tutorial. (Note: In bacteria, the methionine on initiator tRNA^{met} has a formyl group attached to it i.e. it is formylated. The formylated tRNA^{met} prevents the formation of a peptide bond between the amino acids).

Once the initiation complex has been formed a process called **translocation** occurs.

The tRNA recognizes and attaches with its anticodon arm to the correct codon on the mRNA. The P-site of the ribosome is occupied by the initiator tRNA^{met}. The A-site, which is initially empty, is then occupied



G-NETIX

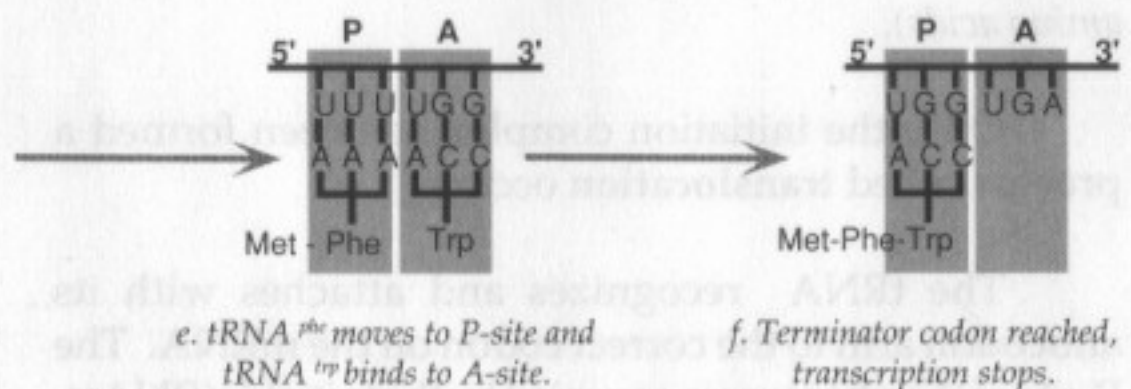


when another tRNA binds with the second codon. In figure 5.7, it is tRNA^{phe}. A peptide bond is formed between the two aminoacids, i.e. between methionine and the phenylalanine. The mRNA then moves three nucleotides relative to the ribosome, so that the codon which is previously in the A-site now moves to P-site. In the example, tRNA^{phe} moves to P-site from A-site. The empty A-site is now occupied by another complementary tRNA. In the example, it is tRNA^{trp}. The process is repeated until a termination codon enters the A-site. The termination codon in the example is UAA. Because there is no tRNA for the termination codon, the synthesis stops.

Termination requires the activity of one of two proteins called **release factors** which binds directly to stop codon. The polypeptide detaches from the final tRNA.

The dissociated ribosome then initiates the translation of another mRNA.

Normally, a strand of mRNA is translated by



several ribosomes at the same time which means multiple copies of the protein are formed almost simultaneously.

When a cell divides it must make a complete copy of its genes in order for each daughter cell to receive complete sets (chromosomes) during the cell cycle in mitosis. A dividing cell, therefore, must make an exact duplication of its DNA and pass it on to the next generation. The duplication of DNA is called DNA replication. Replication occurs during the S phase of the cell cycle.

How does the duplication occur?

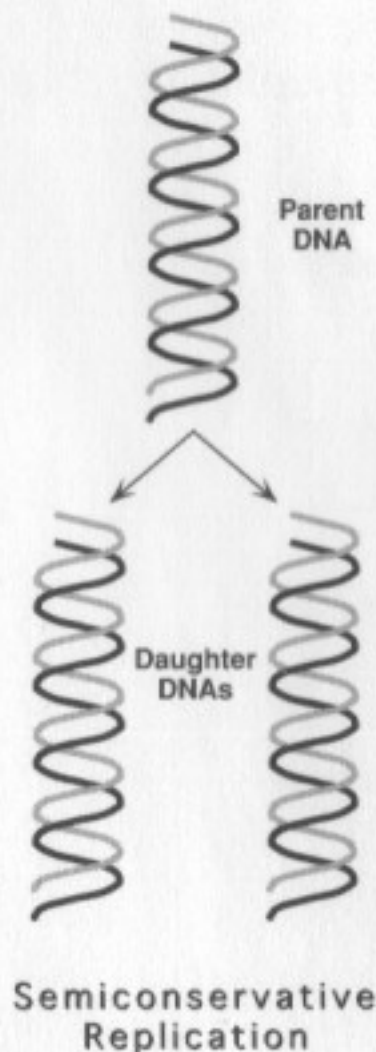
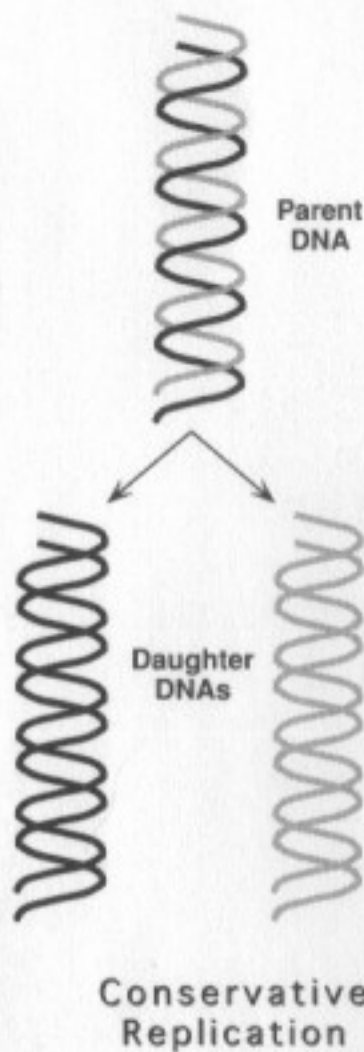
Does each duplex DNA make two new strands and leave the old strands intact (a process called conservative replication)?

Experiments proved that the replication occurs by a semi-conservative process (Figure 4.1) in which each of the two



Chapter 6 DNA Replication

When a cell divides it must make a complete copy of its genes in order for each daughter cell to receive complete sets (*remember the S phase of the cell cycle in mitosis*). A dividing cell, therefore, must make an exact duplication of its DNA and pass it on to the next generation. The duplication of DNA is called **DNA replication**. Replication occurs during the S phase of the cell cycle.



How does the duplication occur?

Does each duplex DNA make two new strands and leave the old strands intact (a process called **conservative replication**)?

Experiments proved that the replication occurs by a **semiconservative process** (figure 6.1). In semiconservative

Figure 6.1: DNA replication

replication, the daughter DNA molecule contains one of the original polynucleotide strands and one newly synthesized strand.

The Replication Process

The replication begins with the separation of DNA strands at a certain point, called the **replication origin**. The separation of the DNA molecule is carried out by enzymes called **helicases**. The breakage gradually progresses in both directions (bidirectional) along the DNA molecule. The part of the DNA molecule where breakage occurs and the new polynucleotides are synthesized is called a **replication fork** (figure 6.2).

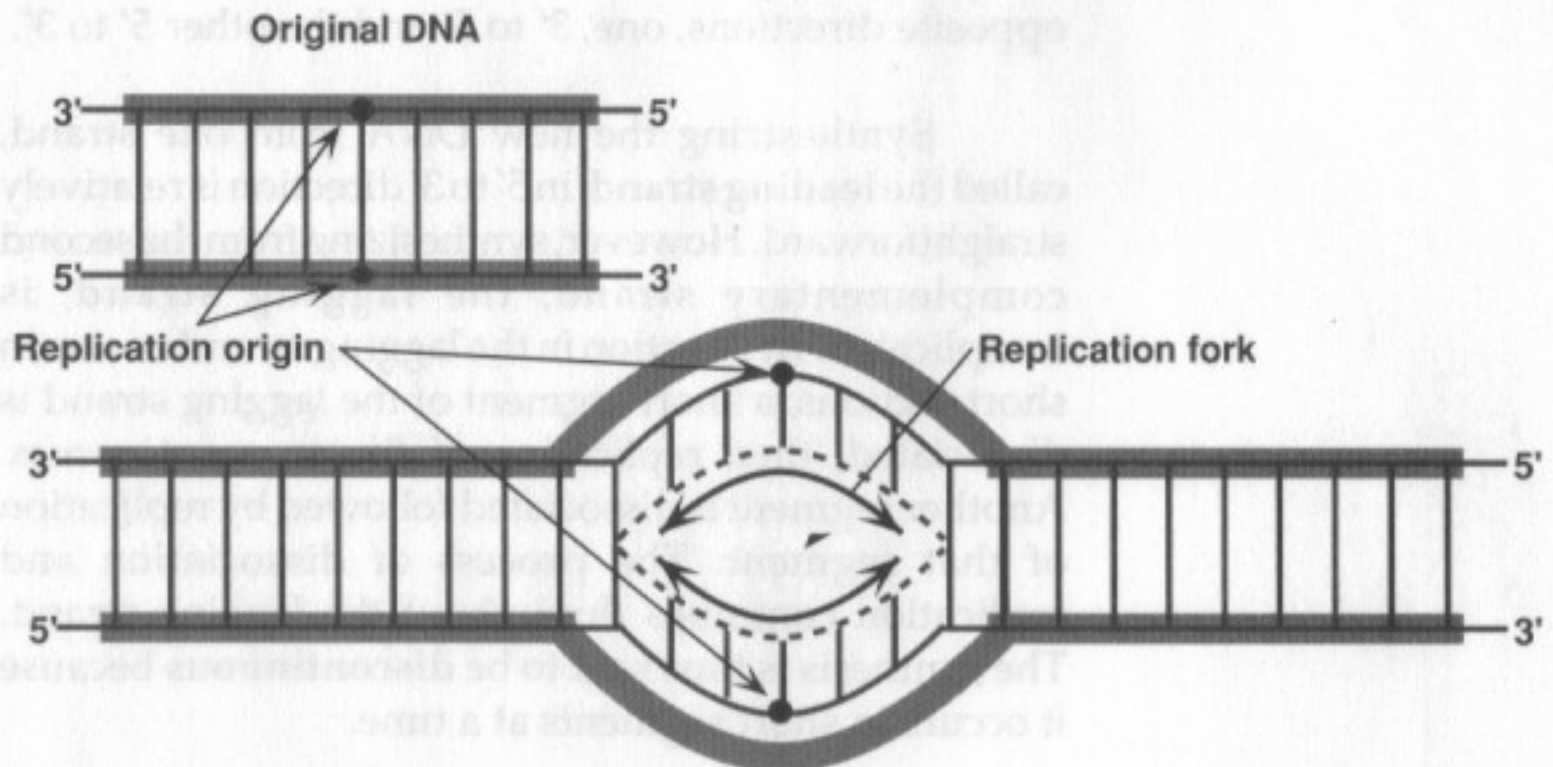


Figure 6.2 Bidirectional replication

DNA replication and synthesis are facilitated by an enzyme called **DNA polymerase**.

The DNA polymerase is identical to RNA polymerase, an enzyme involved in transcription (*see Chapter 5: "How Proteins are Formed"*), except that the DNA polymerase creates a new complementary DNA strand, not RNA.

DNA polymerase synthesizes new DNA only in the 5' to 3' direction. Therefore, the template strand of old DNA is being read in 3' to 5' direction.

The two dissociated strands of the DNA molecule are not synthesized equally. Remember that the two complementary strands of DNA are connected in opposite directions, one, 3' to 5' and the other 5' to 3'.

Synthesizing the new DNA from one strand, called the **leading strand**, in 5' to 3' direction is relatively straightforward. However, synthesizing from the second complementary strand, the **lagging strand**, is complicated. Replication in the lagging strand occurs in short sections; a short segment of the lagging strand is dissociated, then replication of the segment occurs. Another segment is dissociated followed by replication of that segment. The process of dissociation and replication continues throughout the lagging strand. The synthesis is thus said to be **discontinuous** because it occurs in short segments at a time.

The short DNA fragments that are formed are then connected by another enzyme called **ligase** (figure 6.3, pg. 47). *Note: The short DNA fragments are called*

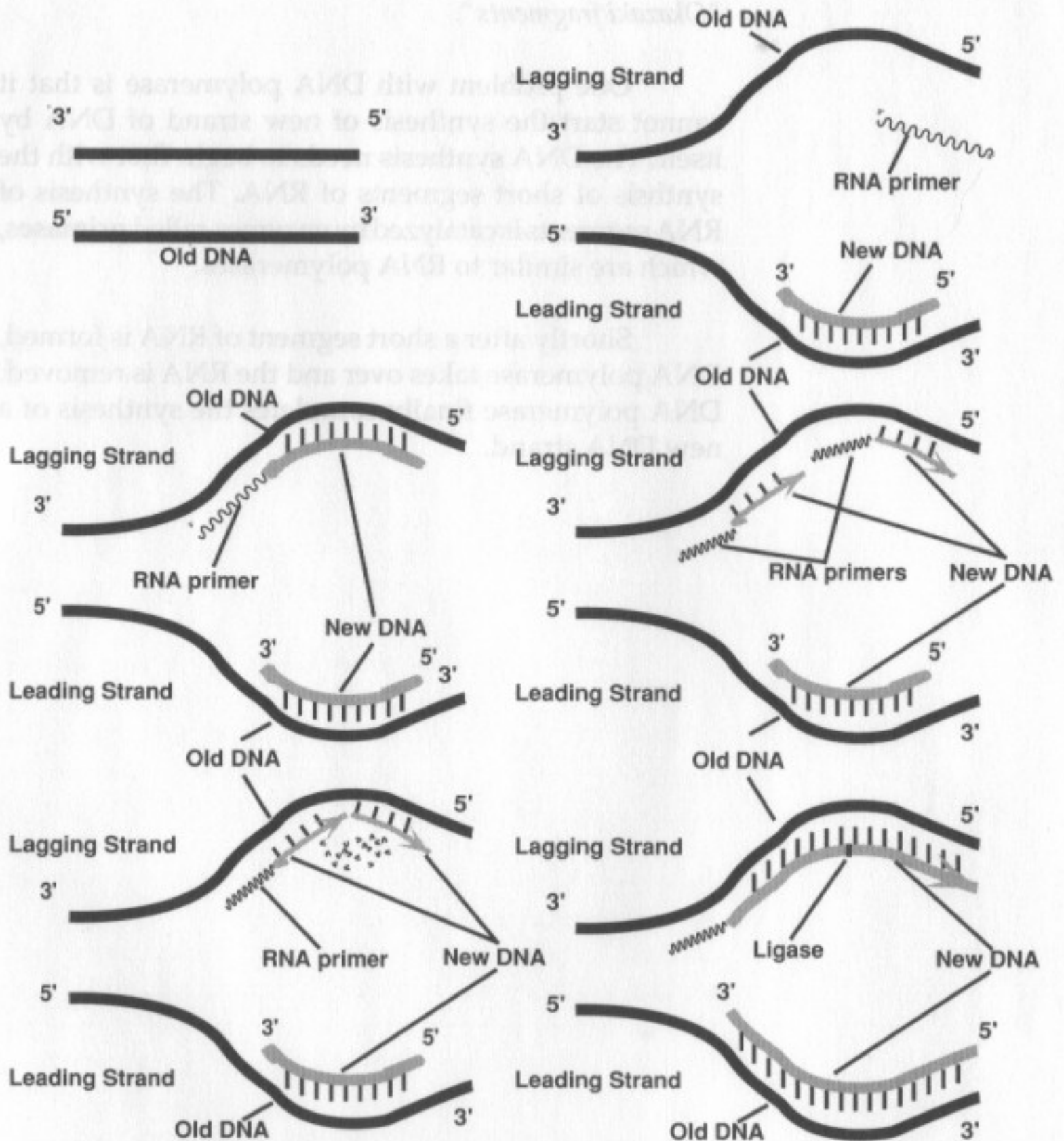


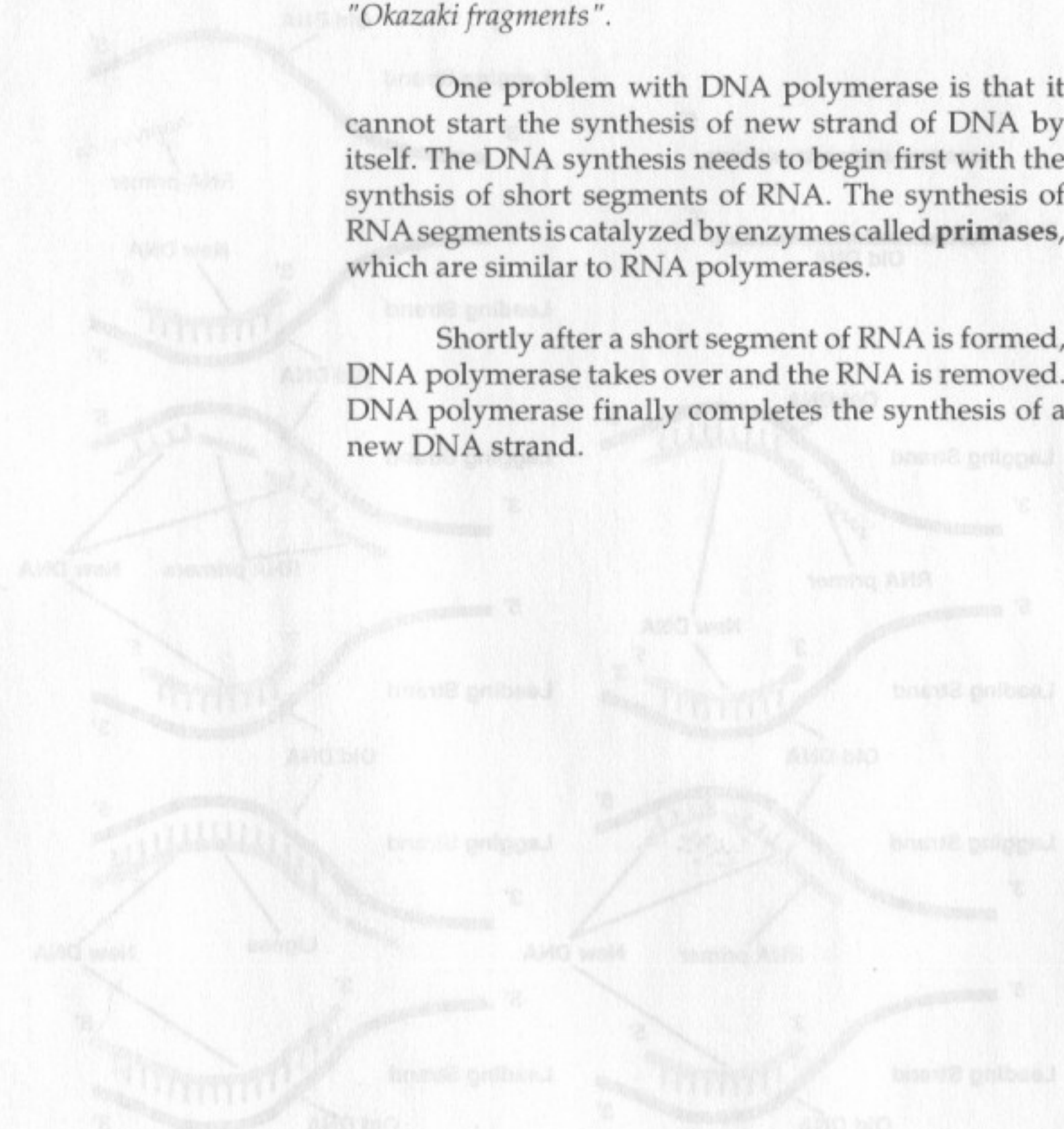
Figure 6.3 Steps in synthesis of leading and lagging strands

G-NETIX

"Okazaki fragments".

One problem with DNA polymerase is that it cannot start the synthesis of new strand of DNA by itself. The DNA synthesis needs to begin first with the synthesis of short segments of RNA. The synthesis of RNA segments is catalyzed by enzymes called **primases**, which are similar to RNA polymerases.

Shortly after a short segment of RNA is formed, DNA polymerase takes over and the RNA is removed. DNA polymerase finally completes the synthesis of a new DNA strand.



Chapter 7

Gene Control and Development

Human beings look very different from each other. Within each of us, the structure and function of one cell type differs from the structure and function of another cell type. Brain cells, for example, look and behave differently from the cells of the heart. Yet all these cells are derived from the same set of genes and all of them are produced from a single cell (zygote) that developed from fusion of the ovum and the sperm.

Why these different appearances occur is not fully known.

Currently, it is thought that cell differentiation occurs not because of different gene sets nor different numbers of genes (all these cells have the same set and same number), but because different genes display their effects at different times. Displaying of these effects is called **expression**. When a gene displays its effect we say **the expression of the gene is turned on** or **the gene is expressed**.

Different gene expressions result in synthesis of different sets of RNA and protein molecules. These proteins then give different appearances and functions to the cells.

Gene expression is regulated. Many genes are turned-on at one time while other genes are turned-off.

Gene expression can be regulated at each step or through a combination of several different steps during the synthesis of protein from the DNA.

For example, gene expression can be regulated at:

1. Transcription.
2. RNA processing (How hnRNA is cut to become mRNA. This is beyond the scope of this tutorial.).
3. Translation.
4. mRNA degradation (whether mRNAs are destroyed before translation occurs). If degradation of mRNA occurs before the translation of the mRNA, there will not be formation of any protein.

Among the steps mentioned above, the regulation of transcription is the most important mode of control of gene expression.

How is gene transcription regulated?

There are many ways the transcription is regulated:

1. Gene expression can be turned-on and turned-off, depending upon changes in the environment.

2. Occasionally, an event causes a set of genes to express. The products of these genes then turn-off the transcription of first set of genes while it may turn-on the transcription of another set of genes. This is called **preprogrammed circuits of gene expression**. This type of control may be seen when there is an infection by a virus.

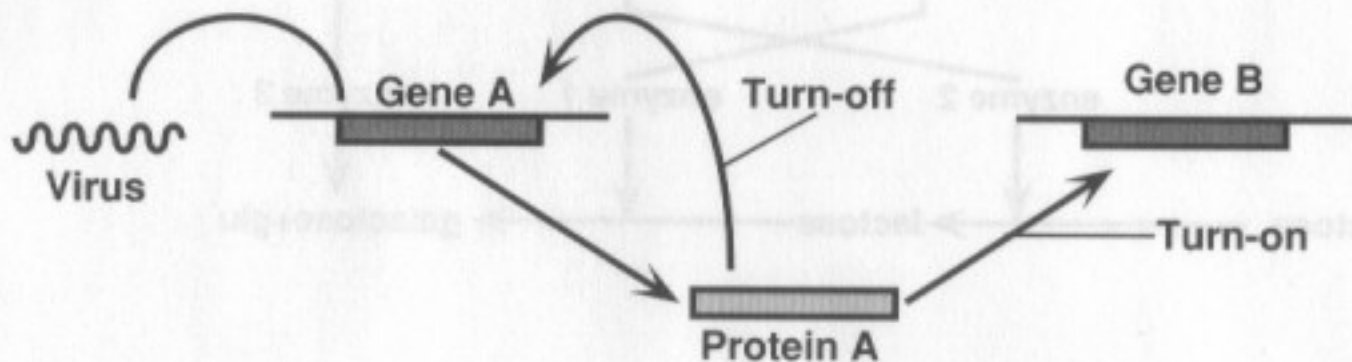


Figure 7.1 Preprogrammed circuits of gene expression

Effect of Changes in the Environment on Gene Expression

The effect on the bacteria (or prokaryote) of changes in the environment is seen in the following example:

In bacteria, there is a set of three enzymes that catalyze the breakdown of a sugar called lactose. Lactose is broken down by the enzymes into two sugars, glucose and galactose. The three enzymes are controlled by a cluster of three genes, *lacZ*, *lacY* and *lacA*. These genes form an operon (refer to chapter four, "Genes"). They are transcribed on the same mRNA and all three have one promoter and one terminator (figure 7.2).

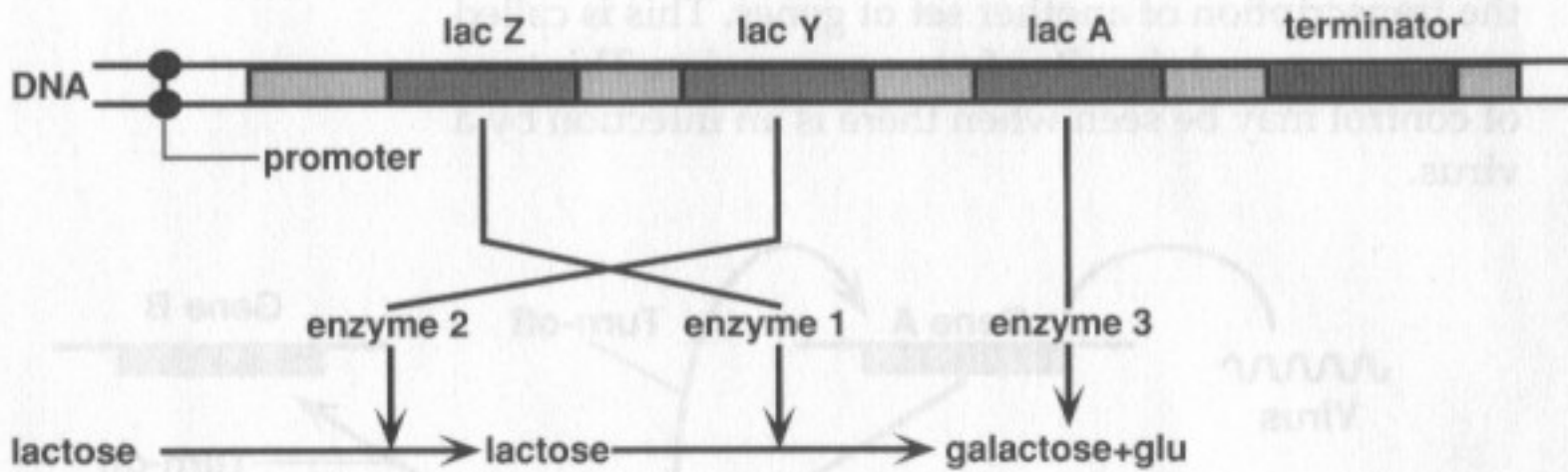


Figure 7.2 lac operon

However, there is an additional gene related to the *lac* operon genes. This gene is called *lacI*. The product of the *lacI* gene regulates the expression of the other three genes (figure 7.3).

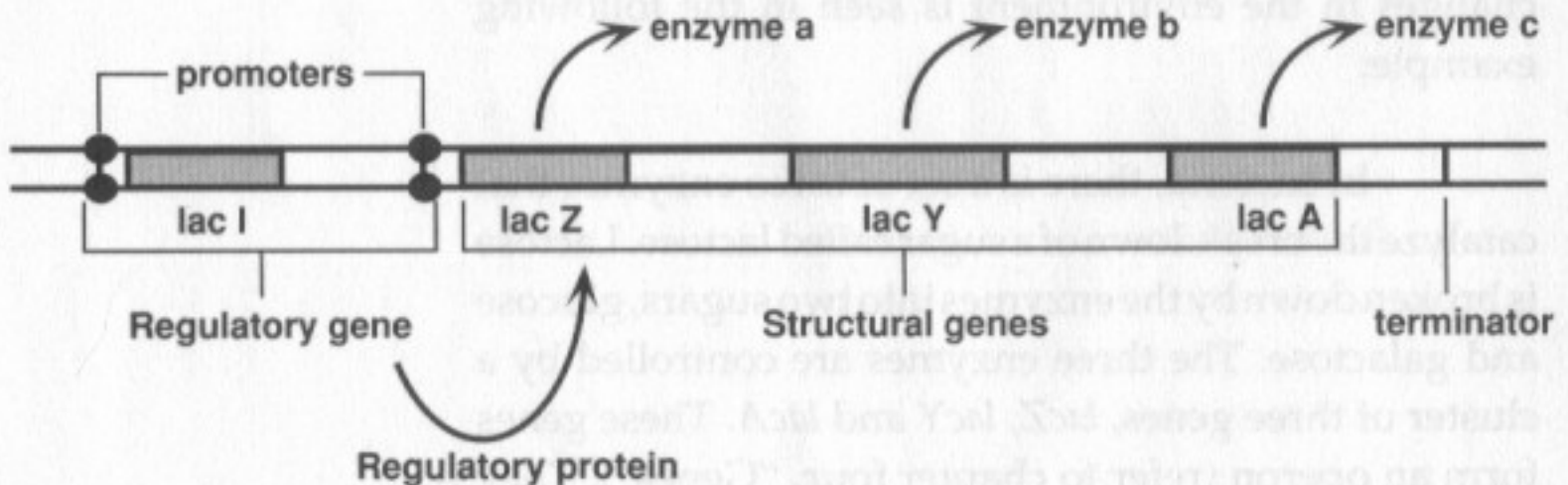


Figure 7.3. The *lacI* gene.

Due to this, *lacI* is called a **regulatory gene** and *lacZ*, *lacY* and *lacA* are called **structural genes**. The product of *lacI* is a protein called **repressor** because it suppresses the expression of the *lac* operon. If there is no lactose in the environment, *lacI* gene of the bacteria produces the repressor protein which binds the *lac* operon at a site between the promoter and *lacZ* gene. The site is called **operator**. The operator overlaps the promoter. When the repressor protein binds the operator it prevents the RNA polymerase from binding the promoter and prevents the transcription of *lac* operon. Because there is no transcription, no enzymes are produced. Therefore, lactose is not broken down to glucose and galactose.

When lactose is present in the environment, however, the lactose binds to the repressor protein itself. This prevents the repressor protein from binding the operator of the *lac* operon gene. There is no regulation of transcription of *lac* operon. The RNA polymerase binds the promoter and proceeds with the transcription of *lac* operon to form the enzymes. The enzymes then split lactose into glucose and galactose. Thus transcription of *lac* operon occurs when lactose is present. When lactose is absent there is no transcription of *lac* operon. The process where the expression of genes is turned on because of a presence of a substance in the environment is called **induction**. When expression is regulated in this manner genes are referred to as **inducible genes** (see figure 7.4, pg. 54).

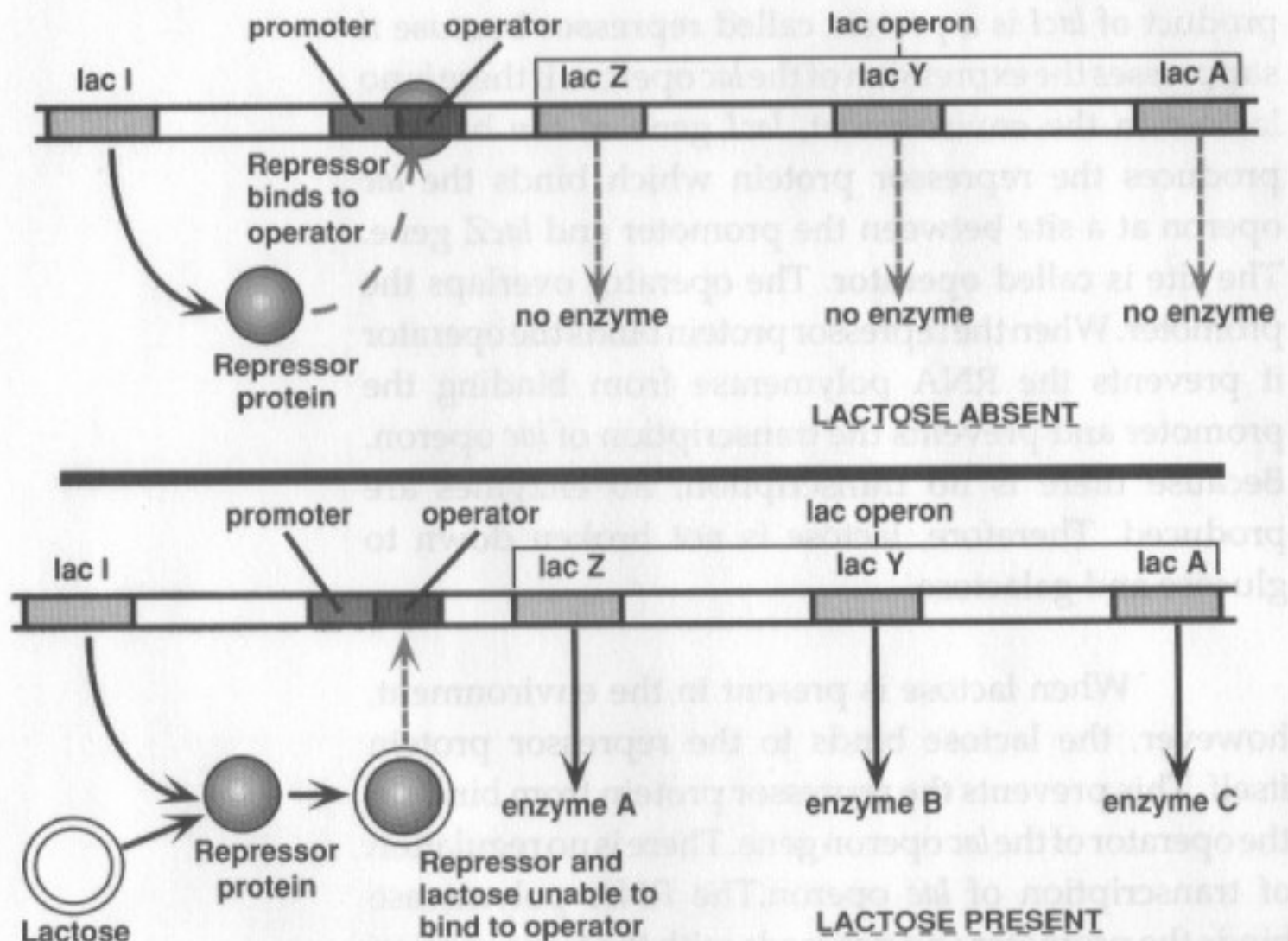


Figure 7.4 Regulation of lac operon when lactose is absent and when lactose is present

A similar but different type of regulation is also present in the bacteria. It is related to an amino acid named **tryptophan**.

The bacteria has a tryptophan operon gene which allows the bacteria to synthesize its own tryptophan even though not very effectively. The tryptophan operon produces enzymes that catalyse the synthesis of this amino acid. However, when it is present

in the environment the bacteria stops synthesizing its own tryptophan. The tryptophan binds the repressor protein and the **tryptophan-repressor complex binds to the operator of the tryptophan operon**. This prevents the RNA polymerase from binding the promoter and prevents the transcription of tryptophan operon. As a result no tryptophan is synthesized (figure 7.5).

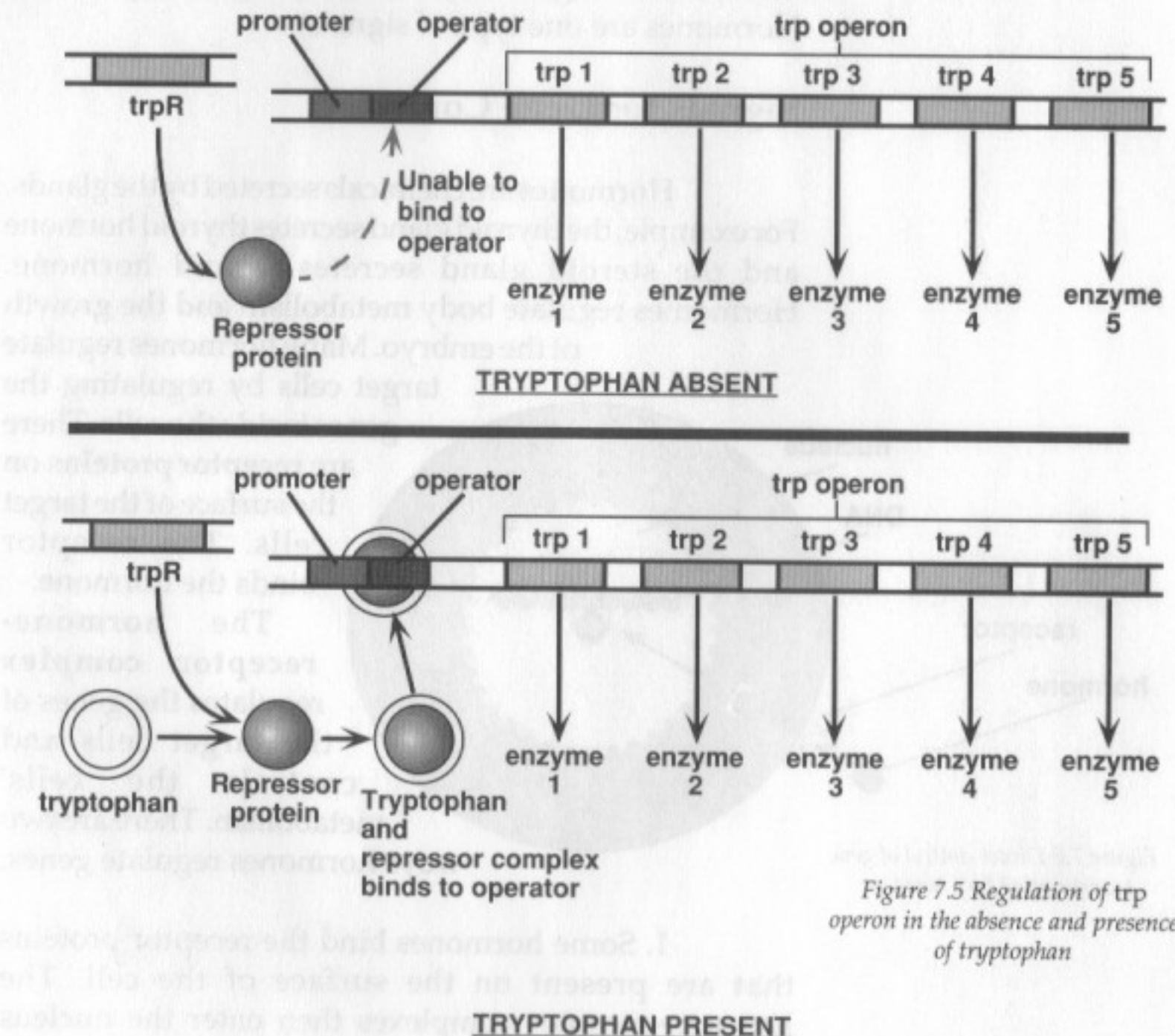


Figure 7.5 Regulation of *trp* operon in the absence and presence of tryptophan

Regulation of gene expression in Human and Eukaryotes

In humans and eukaryotes the control of gene expression is more complicated. Many of the processes are not well known. However, evidence have shown that there are signals that control the gene expression. Hormones are one type of signals

Signals for Gene Control

Hormones are chemicals secreted by the glands. For example, the thyroid gland secretes thyroid hormone and the steroid gland secretes steroid hormone. Hormones regulate body metabolism and the growth of the embryo. Many hormones regulate target cells by regulating the genes inside the cells. There

are **receptor proteins** on the surface of the target cells. The receptor binds the hormone.

The **hormone-receptor complex** regulates the genes of the target cells and controls the cells' metabolism. There are two ways hormones regulate genes:

1. Some hormones bind the receptor proteins that are present on the surface of the cell. The hormone-receptor complexes then enter the nucleus

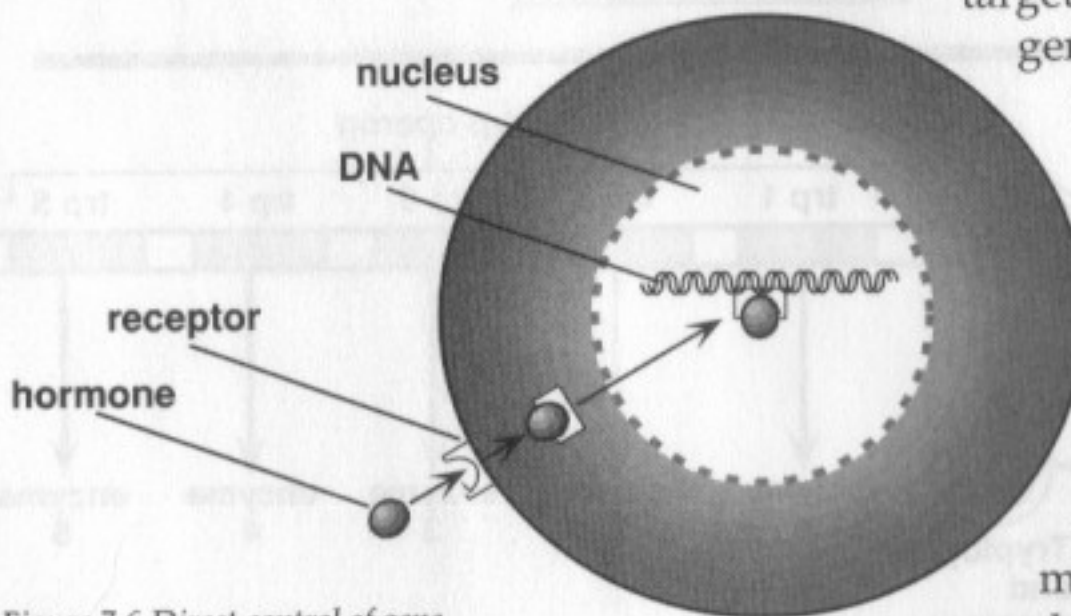


Figure 7.6 Direct control of gene transcription by a hormone.

and directly regulate the genes by controlling the transcription of the genes (figure 7.6).

2. In contrast, some hormones bind to specific cell-surface receptors and exert chain of reactions inside the cells without entering the cells (figure 7.7).

In addition to hormones there are other proteins that are released from one cell type and exert dramatic effects on other neighboring or distant cell types. These are called **growth factors**. For example, **nerve growth factor** causes nerve cells to grow and **epidermal growth factor** speeds up development of limbs, skin, and integuments.

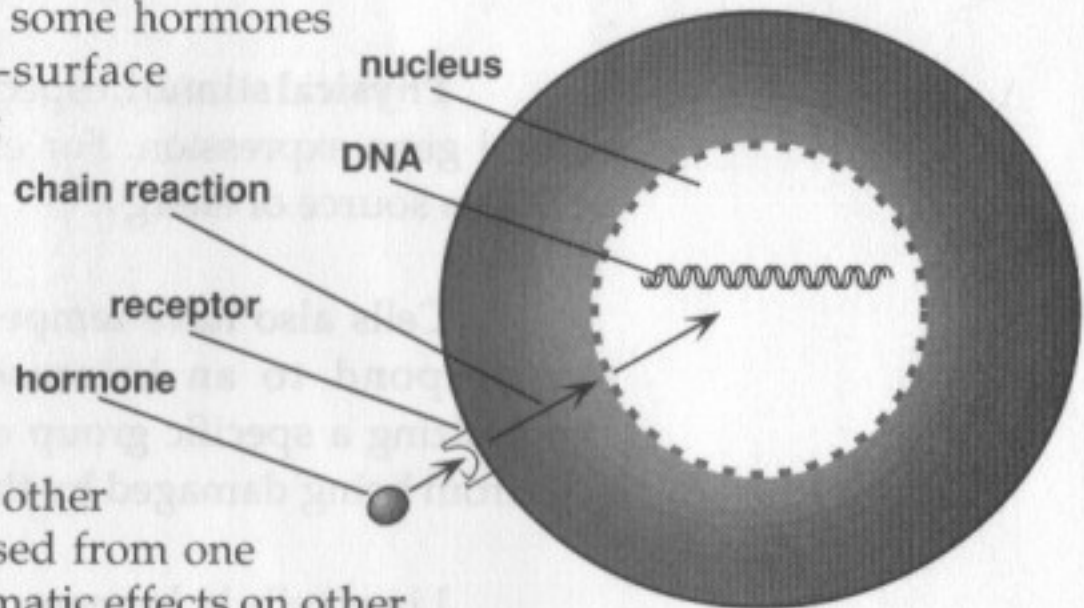


Figure 7.7 Hormone-receptor complex causing chain of reaction

There is yet another mechanism for regulating the growth of the cells. This is called **Cell-Cell contact**. For the cells to develop, two or more different cell types must make contact with one another. For example, in the early-stage embryo there are three layers of cells, **ectoderm**, **mesoderm** and **endoderm**. It is found that the endoderm cells bud out at different areas and touch the cells of the mesoderm. This interaction results in the subsequent development of salivary glands, lungs, and the liver.

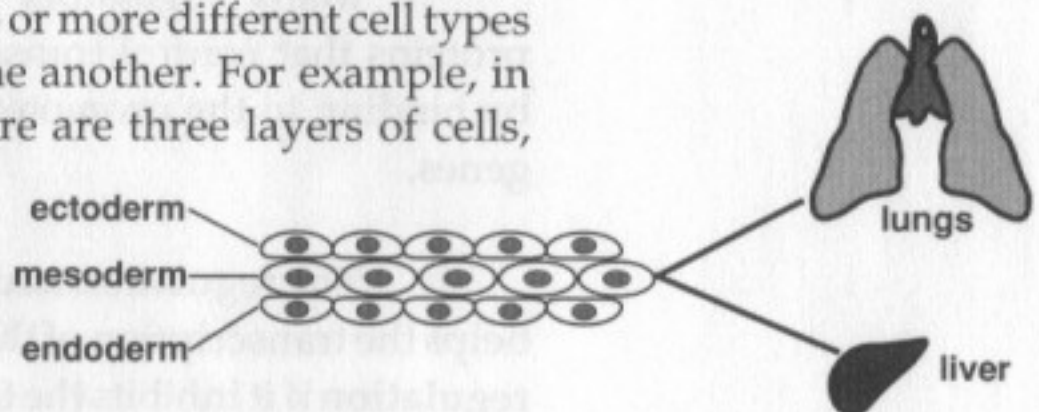


Figure 7.8. Cell-cell contact.

G-NETIX

Environmental and **nutritional signals** can also control gene expression. This is more common in bacteria and smaller eukaryotes such as yeast and molds.

Physical stimuli, especially heat and light, also control gene expression. For example, plants require light as a source of energy.

Cells also have **temperature-sensitive genes** that respond to an increase in temperature by synthesizing a specific group of proteins that protect cells from being damaged by the sudden intense heat.

Liver cells in humans and animals are able to detoxify noxious chemicals and other toxic substances by producing *metallothioneins*, proteins that bind heavy-metal ions. These proteins bind the heavy metals and prevent the metals from damaging the organs and cells.

Many regulatory genes produce regulatory proteins that control transcription of structural genes by binding to the promoter sequence of the structural genes.

The regulation may be **positive regulation** if it helps the transcription of the adjacent gene, or **negative regulation** if it inhibits the transcription. More than one regulatory protein may act on the structural gene and different regulatory proteins may interact with one another.

For example, regulatory protein A may turn a set of genes into a muscle cell. Together with regulatory protein B, it may turn a different set of genes into a liver cell. A perfect example, is steroid hormone. Steroid hormone produces different reactions in different cells.

Not all regulatory proteins are equal. There are **master gene regulatory proteins** that control other genes. For example, from the experiments performed on a fly named *Drosophila*, mutation of one gene converts one part of the fly's body to another. One mutation causes the adult fly to have a leg growing out of its head instead of an antenna.

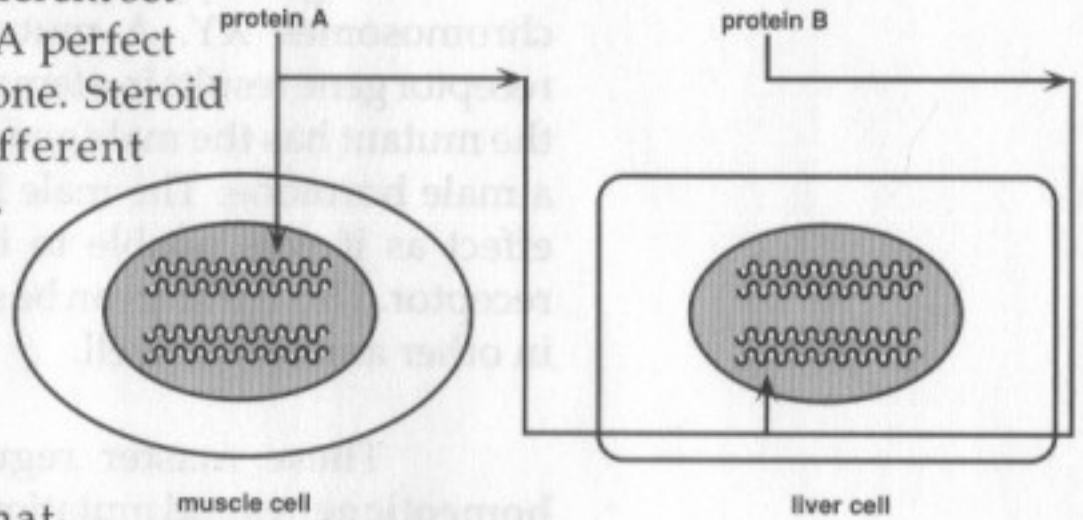


Figure 7.9 Regulatory protein A turns different sets of genes in different cells

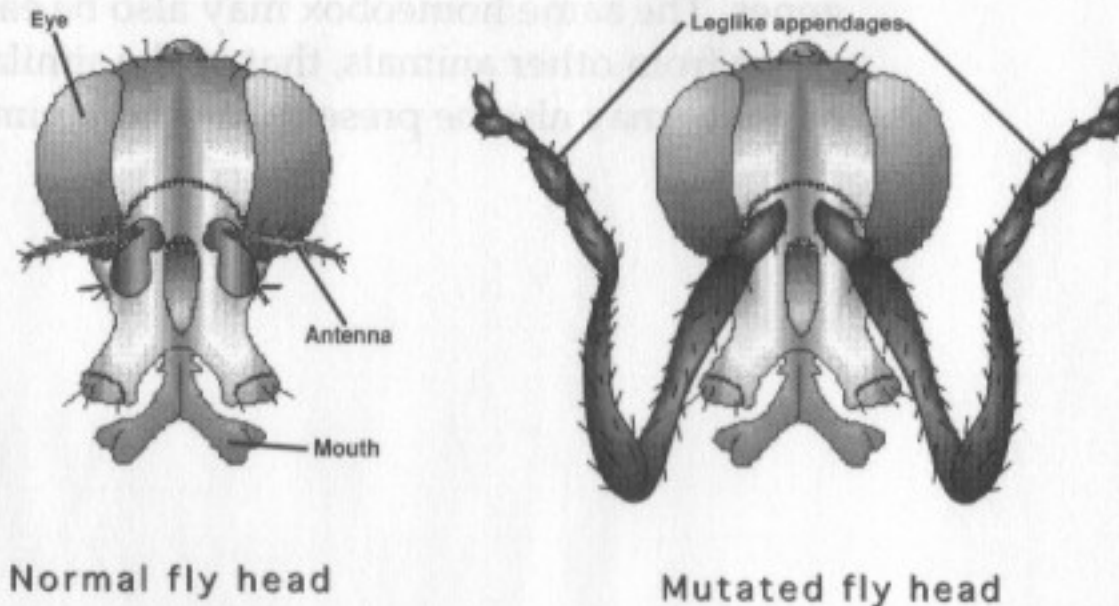


Figure 7.9 Mutation in *Drosophila*

Another example is in human. Normal man has a male genotype (gene composition) with the sex chromosomes XY. A mutation of a male hormone receptor gene results in a female appearance even though the mutant has the male genotype as well as producing a male hormone. The male hormone can not show its effect as it was unable to interact with the mutated receptor. This feature can be seen not just in humans but in other animals as well.

These master regulatory genes are called **homeotic genes** and mutations that bring about changes in the body structures are called **homeotic mutations**.

In the homeotic genes a sequence of 180 nucleotide pairs, called **homeobox** is present. The unique feature of the homeobox is that the gene sequence is the same except for a few minor variations in all the homeotic genes. The same homeobox may also be easily seen in genes from other animals, that is, the similar sequence of genes may also be present in other animals.

Chapter 8 **Mutation**

During the evolution of species alterations in genetic material occur from **mutation** and **recombination**. This allows the species to be able to adapt to its new environment.

Mutations are changes which occur in the nucleotide sequence of a DNA.

Mutation may or may not result in a change in the **Phenotype**. Phenotype is the appearance of the organism. A man and a gorilla have different appearances. Therefore, they have different phenotypes.

Genotype is the genetic structure of an organism. Monkeys and chimpanzees may look alike (similar phenotypes) but they may have different gene composition (different genotypes).

The organism displaying a new appearance as a result of mutation is called a **mutant**. For example, the *Drosophila* fly, with leglike appendages, is a mutant.

Mutation provides the basis for evolution. Without mutations organisms would not be able to evolve and adapt to the changing environment.

However, mutations can also cause undesirable side effects. Some mutations may even kill the organism.

Mutations may occur spontaneously or be caused by insulting agents.

In a bacteria, mutation is said to occur spontaneously once in every 10^8 cell divisions. In man it is probably higher. Most of the mutations occur at random.

Mutations can also be induced by exposure to agents such as **ionizing irradiation**, **ultraviolet light**, and **chemical agents**.

Agents that can cause mutations are called **mutagens**. Mutations can also result from infection with a virus.

There are different types of mutations:

1. **Point mutation** is said to occur when one nucleotide is replaced by another.

C G **G** C T G \longrightarrow C G **C** C T G

2. **Insertion** inserts one or more nucleotides to the gene.

C G G C T G \longrightarrow C G G **T** C T G

3. **Deletion** deletes one or more nucleotides from the gene.

C G **G** C T G \longrightarrow C G C T G

4. **Inversion** excises a portion of the gene and reinserts it at the same position in reverse.

C G **G C** T G → C G **C G** T G

Sometimes, a second mutation may occur in the same gene and restore the original amino acid sequence. This is called **back-mutation**. When back-mutation occurs the amino acid sequence does not change even though the genetic code is different (Remember, some amino acids are encoded by two or three different genetic codes).

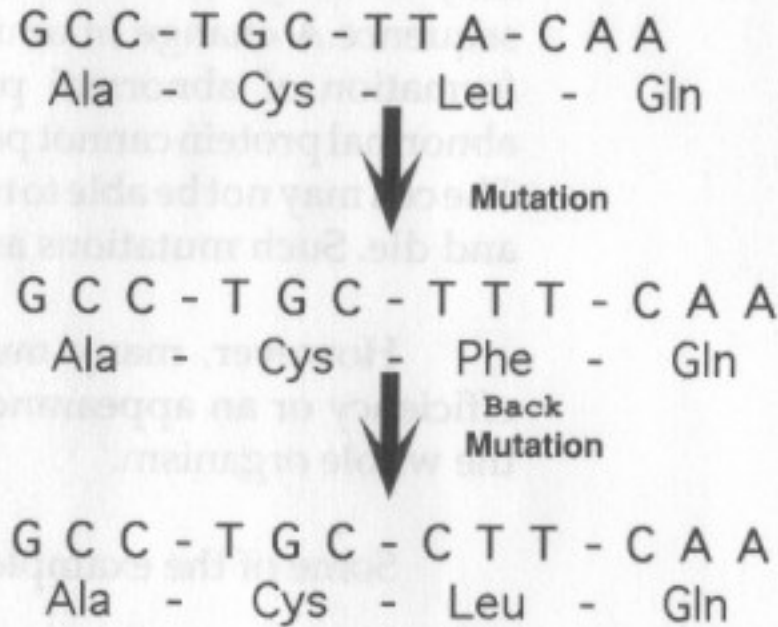
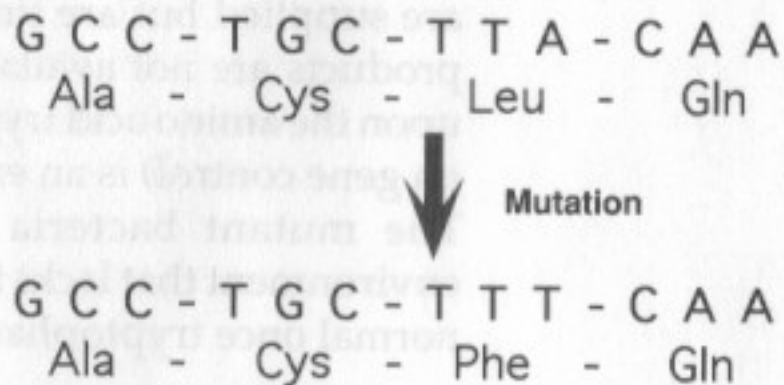


Figure 8.1. Back mutation.

In some cases, a mutation may occur in a different gene which, in turn, suppresses the mutated first gene. This is called a **suppressor mutation**.



mutation in
another gene

Figure 8.2 Suppressor mutation.

If a point mutation occurs in the intergenic region (intron), the region which is not the genetic code, nothing will happen.

If a mutation occurs in a genetic code, misinterpretation of the genetic code may occur. This may or may not result in a change of amino acid sequence. A change in amino acid sequence results in formation of abnormal protein or polypeptide. The abnormal protein cannot perform its function in the cell. The cell may not be able to tolerate the loss of its function and die. Such mutations are called **lethal mutations**.

However, many mutations may just affect the efficiency or an appearance (phenotype) of the cell or the whole organism.

Some of the examples are:

Auxotrophic mutants are mutants that are unable to synthesize essential products by themselves. These mutants can survive and grow if the essential products are supplied but are unable to survive if the essential products are not available. The bacteria that depends upon the amino acid tryptophan (refer to chapter seven on gene control) is an example of auxotrophic mutant. The mutant bacteria cannot survive long in an environment that lacks tryptophan. The bacteria grows normal once tryptophan is available.

Temperature Sensitive mutants are able to survive at one temperature range. The mutants die if the temperature is raised above the range.

There are different mutation effects on the amino acid sequence.

1. **Silent mutation** occurs when the mutation changes the nucleotide in the codon but does not change the amino acid.



2. **Missense mutation** occurs when the mutation changes the nucleotide in the codon resulting in a different amino acid.



3. **Frameshift mutation** occurs when there is an insertion or deletion of one or more of the nucleotides.



Chapter 9 **Chromosomes**

In a human cell chromosomes, with the exception of **sex chromosomes**, are present in two exact copies (**diploid**). One copy of chromosomes is inherited from the father and the other copy from the mother.

Sex chromosomes are labeled X and Y. They are called **sex chromosomes** because Y chromosome is present only in male and X chromosome is present in both male and female. In male, the Y chromosome is inherited from the father and X from the mother.

The other chromosomes are called **autosomes**.

There are altogether 23 pairs of chromosomes in humans, i.e 46 chromosomes: 22 pairs of autosomes and one pair of sex chromosomes. The chromosome composition of a man is therefore 22 + XY and the chromosome composition of a woman is 22 + XX.

The number of chromosome pairs differs in different species. A gorilla has 24 pairs of chromosomes and a chicken has 39 pairs. If there is only one chromosome, as in the case of sperm and ovum (reproductive cells), it is called **haploid**.

Each DNA molecule is packaged in a separate chromosome. Each chromosome contains from 50×10^6 to 250×10^6 nucleotide pairs.

Chromosomes are made of protein and the DNA genes. Figure 9.1 is a structure of a chromosome under different magnifications.

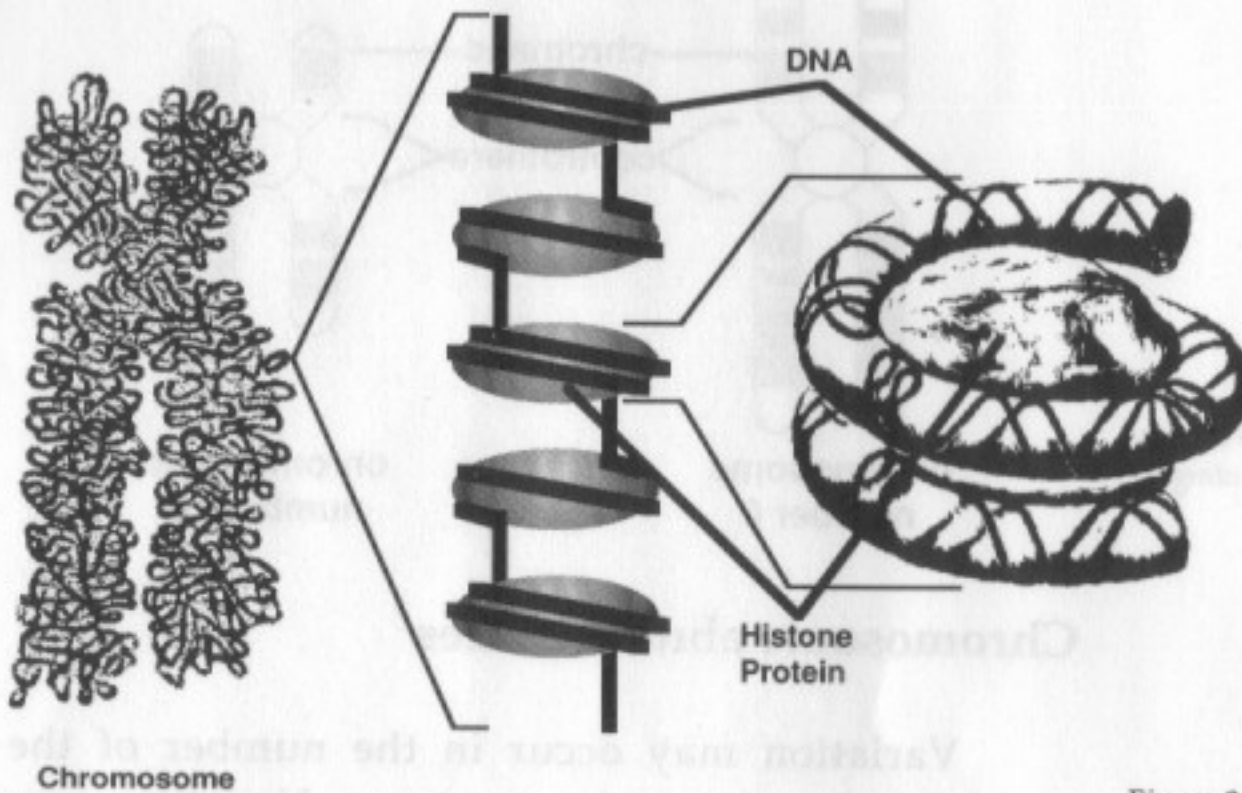


Figure 9.1 Chromosome structure under different magnifications

Individual chromosomes are not identifiable except during cell division. At the metaphase of cell division the chromosomes become visible.

By using techniques called **chromosome banding**, the chromosome pair is displayed during the metaphase as two identical strands called **chromatids** that are joined together by a structure called **centromere**. The position of the centromere is unique for each pair and is used to identify them (see chromosome map). Areas of chromosomes that appear lighter or darker than the

adjacent regions when treated with particular staining methods appear as bands (figure 9.2).

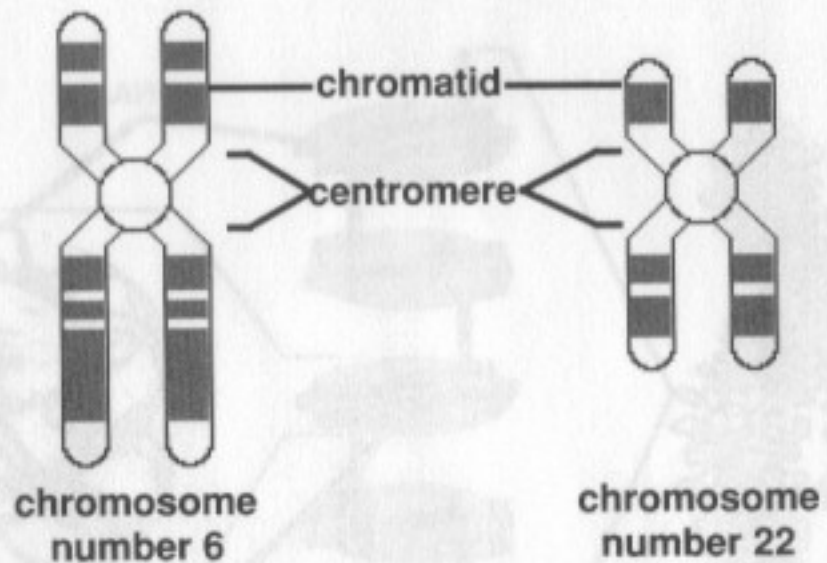


Figure 9.2 Chromosomes as seen with chromosomal banding techniques

Chromosome abnormalities

Variation may occur in the number of the chromosomes. Variation may also occur in the structures of the chromosomes. These variations may result in abnormal appearances and defects in the phenotype (appearance) of the organism.

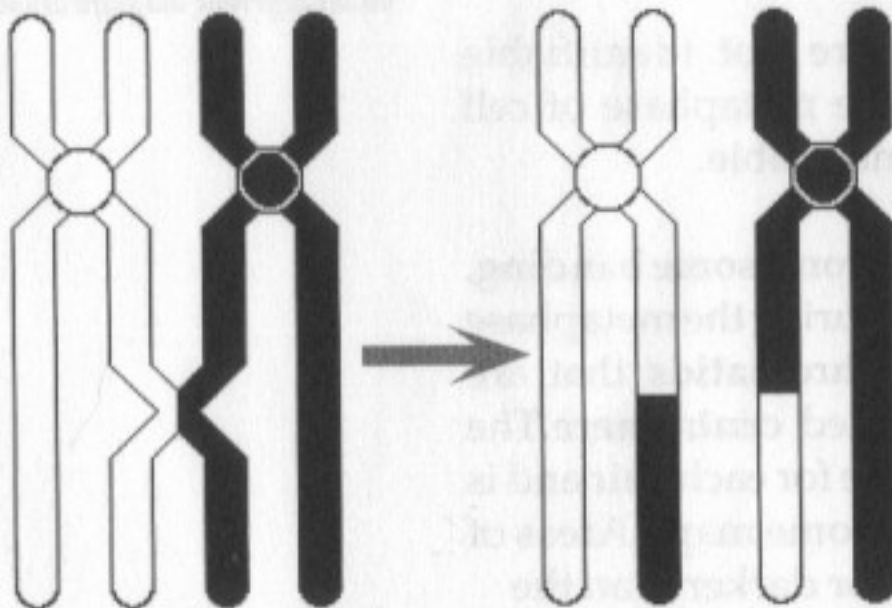


Figure 9.3 Recombination between the chromatids

Variations occur because of **recombination**. Recombination of genes occurs when parts of the chromosomes become

entangled during DNA replication and cell division. The entangled parts switched between the two chromosomes. This results in a new combination of genes not seen in the parent chromosomes (figure 9.3).

Changes in the structure of the chromosomes require breaks in the chromosomes.

There are four different kinds of structural changes:

1. **Deficiencies**, where parts of chromosomes are lost or deleted (figure 9.4 b).

2. **Duplications**, where parts of chromosomes are added or duplicated (figure 9.4c).

3. **Inversions**, where parts of chromosomes become detached and reunited in reverse order (figure 9.4 d).

4. **Translocations**, where parts of chromosomes become detached and joined to different chromosomes (figure 9.5, pg. 70).

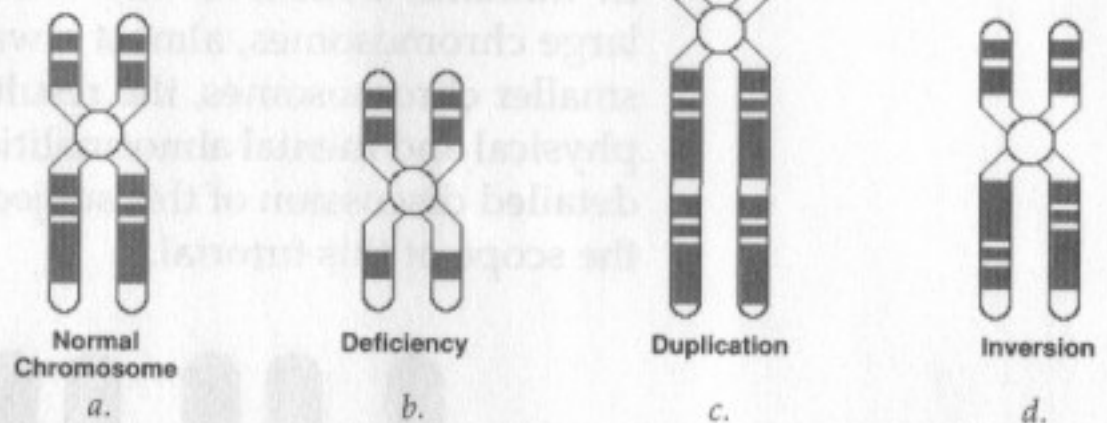


Figure 9.4. Abnormalities in chromosomes.

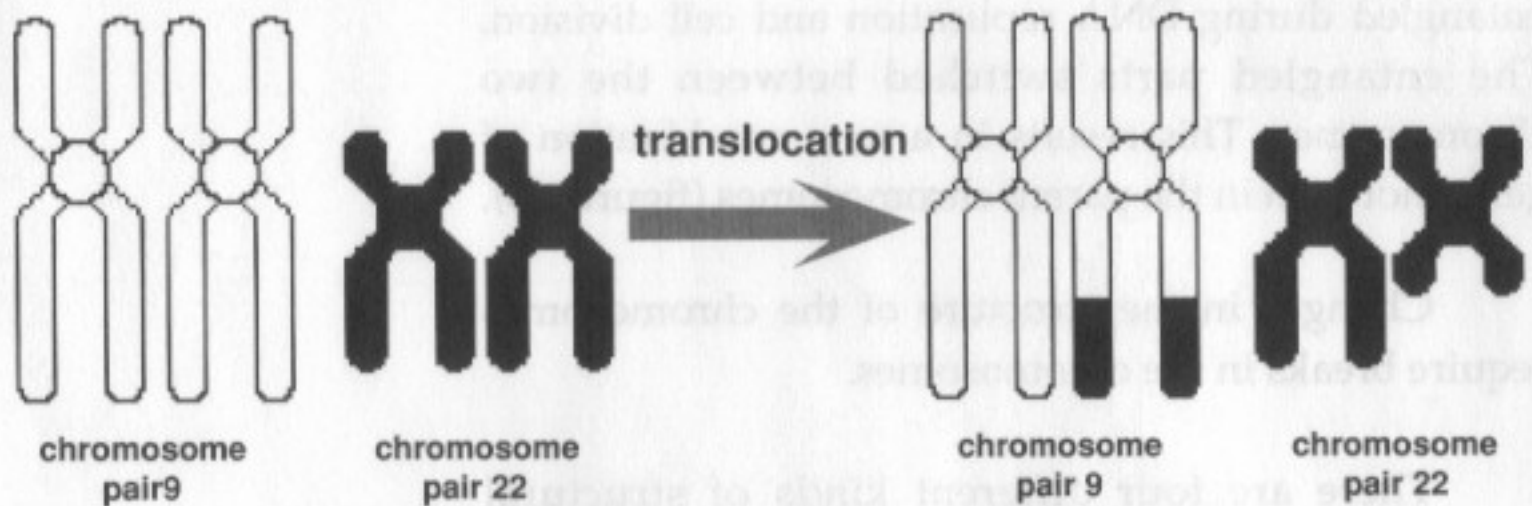


Figure 9.5 Translocation between chromosomes 9 and 22

Variations in number of chromosomes occur when there are deletions of whole or sets of chromosomes or additions of extra chromosomes.

In humans, additions and deletions, particularly of large chromosomes, almost always result in death. In smaller chromosomes, the results may show multiple physical and mental abnormalities (figure 9.6). A more detailed discussion of this subject, however, is beyond the scope of this tutorial.

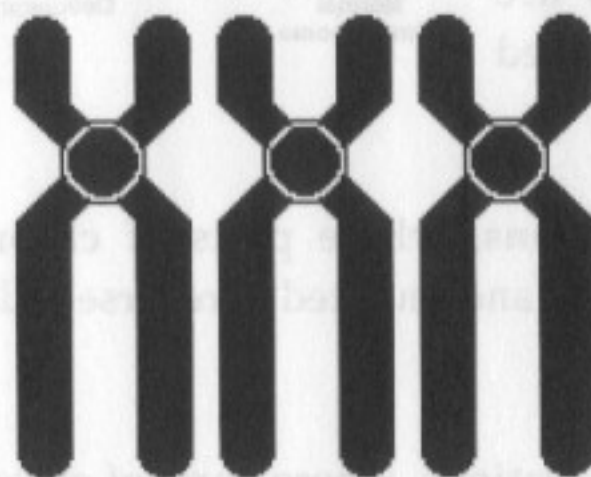


Figure 9.5 Chromosome 21 has an extra chromosome (This is called trisomy)

Chapter 10

Mendel's Findings

Discussing genetics without discussing Gregor Mendel's experiments and findings would be a sin.

He was one of the early pioneers of genetics and his contribution to the field cannot be overemphasized.

Mendel was an Austrian monk who performed experiments with the garden pea in a monastery garden.

He chose peas for his experiments because they have flowers that contain both female and male parts. They also have the capability of **self-fertilization**. Self-fertilization is a process by which the pollens from a plant fertilize the ovules of the same plant.

Garden peas also fertilize with each other easily. This is called **cross-fertilization**.

These pea plants have two alternative forms: plants with tall stems and plants with short stems.

Mendel first experimented by cross-fertilizing a tall-stemmed plant with a short-stemmed plant. The first generation (F1-generation) plants that grew from this cross fertilization were found to be all tall-stemmed. He then self-fertilized the first generation plants (F1-generation).

Second generation plants grew in the approximate ratio of three tall-stemmed plants to one short-stemmed plant (there were 787 tall stems and 277 short stems (figure 10.1).

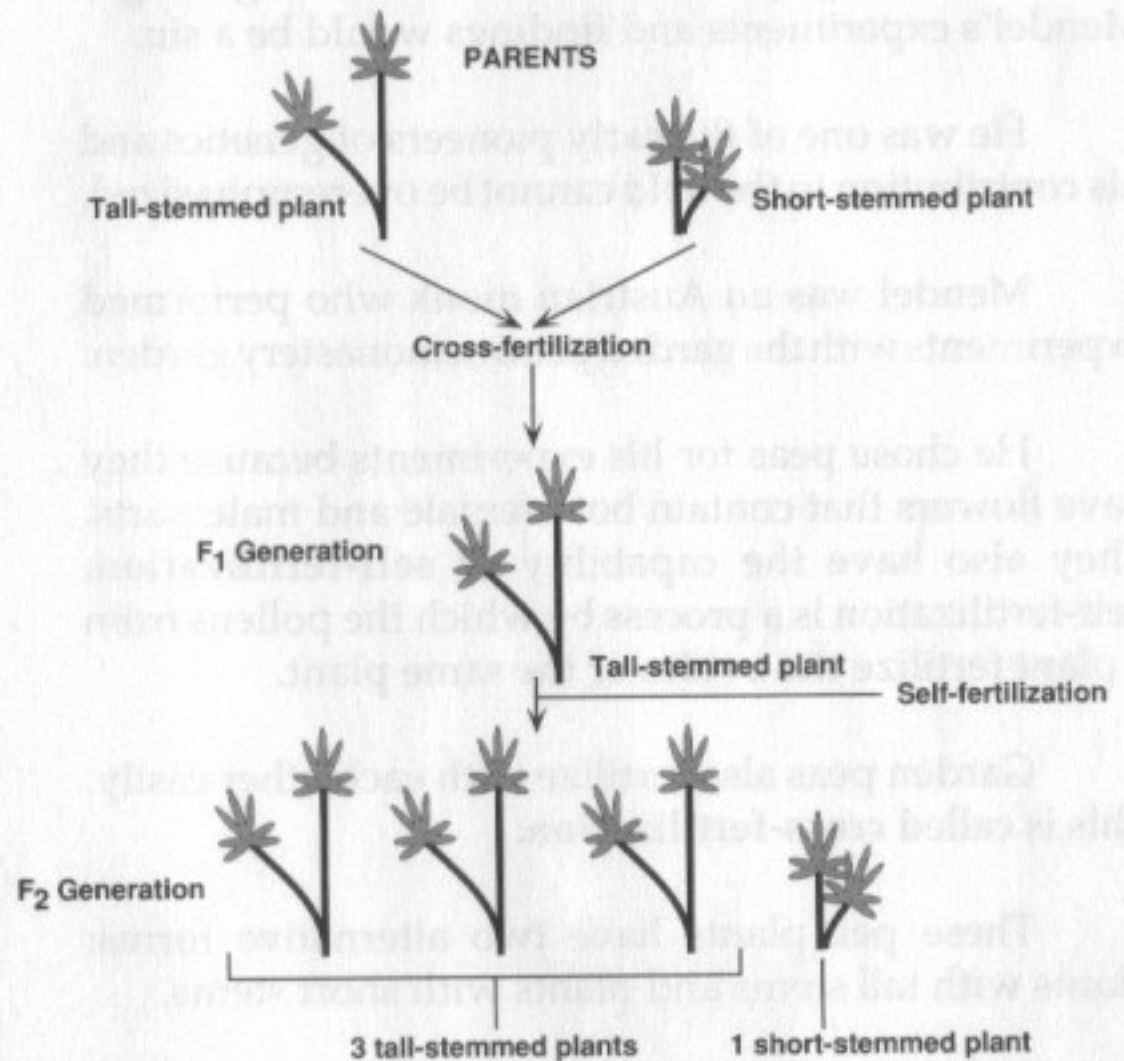


Figure 10.1. Mendel's experiment.

Next, he experimented with the color of the flower. He chose a plant with a violet-colored flower and a plant with a white-colored flower. He noted that after cross-fertilization all the first generation (F₁-generation) plants had violet flowers. After self-fertilization of the

F1 generation, the F2 generation plants produced flowers with approximate ratio of three violet-colored flower plants to one white-colored flower plant.

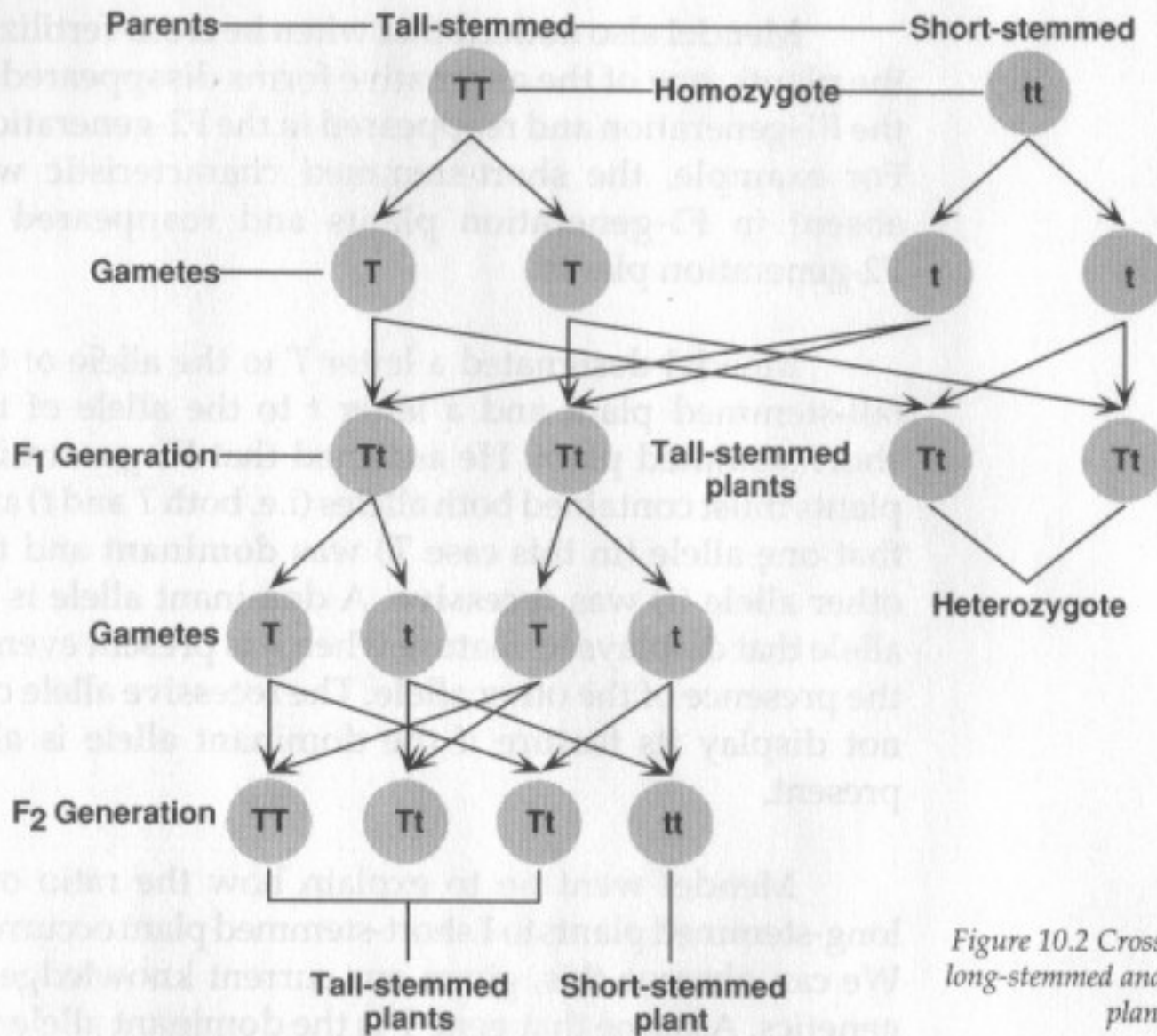


Figure 10.2 Cross-fertilization of long-stemmed and short-stemmed plants.

After observing these findings, Mendel proposed that each characteristic, i.e. stem length or flower color, was a **unit factor**, and that each factor could exist in more than one form. (the plants could exist with either long stems or with short stems). We now understand

that the unit factors Mendel proposed are the **genes** of the plant and the alternative forms of genes are called **alleles**. The color of the flower is a unit factor (gene) and that violet color and white color are alleles.

Mendel also noticed that when he cross-fertilized the plants, one of the alternative forms disappeared in the F1-generation and reappeared in the F2-generation. For example, the short-stemmed characteristic was absent in F1-generation plants and reappeared in F2-generation plants.

Mendel designated a letter T to the allele of the tall-stemmed plant and a letter t to the allele of the short-stemmed plant. He assumed that F1-generation plants must contain both alleles (i.e. both T and t) and that one allele (in this case T) was **dominant** and the other allele (t) was **recessive**. A dominant allele is an allele that displays its feature when it is present even in the presence of the other allele. The recessive allele can not display its feature if the dominant allele is also present.

Mendel went on to explain how the ratio of 3 long-stemmed plants to 1 short-stemmed plant occurred. We can observe this, given our current knowledge of genetics. Assume that gene T is the dominant allele for the stem-length gene and gene t is the recessive allele. F1-generation plants have the genotype Tt .

The plants that have the genotype Tt are referred to as **heterozygotes**, since they contain both alternative alleles.

The plants that contain the genotype TT or tt are referred to as **homozygotes**. The homozygote contains only one form of allele either the dominant allele or the recessive allele.

The alleles of each parent separate during reproduction to produce gametes (in humans, they are egg and sperm). Each gamete contains just one allele (either T or t). Each gamete from tall-stemmed plant TT has only T allele and each gamete from short-stemmed plant has only t allele. When the gamete from one parent fuses with the gamete from the other parent, the resulting daughter contains a pair of alleles. Depending upon the type of alleles inherited, the daughter may be a heterozygote (Tt) or a homozygote (tt). For example, when the gamete T fuses with the gamete t , the F_1 generation plants contain Tt and since T is dominant, the F_1 -generation plants will all have tall-stems.

When the F_1 -generation plant produces the gametes. One gamete has T allele and the other has t allele. When the gametes T and t of one F_1 generation plant fuse with the gametes T and t of another F_1 -generation plant, there are four possible combinations in F_2 -generation plants; TT , Tt , tT , and tt . Because T allele is dominant, there will be three tall-stemmed plants and one short-stemmed plant.

Mendel experimented further with two sets of characteristics. He used plants with Smooth (SS), yellow peas (YY) and plants with wrinkled (ss), green (yy) peas. Smooth surface and yellow color are dominant alleles

while wrinkled surface and green color are recessive alleles. He noted that the resulting F₂-generation showed 9 plants with smooth, yellow peas, 3 plants with smooth, green peas, 3 plants with wrinkled, yellow peas, and 1 plant with wrinkled, green pea (figure 10.3).

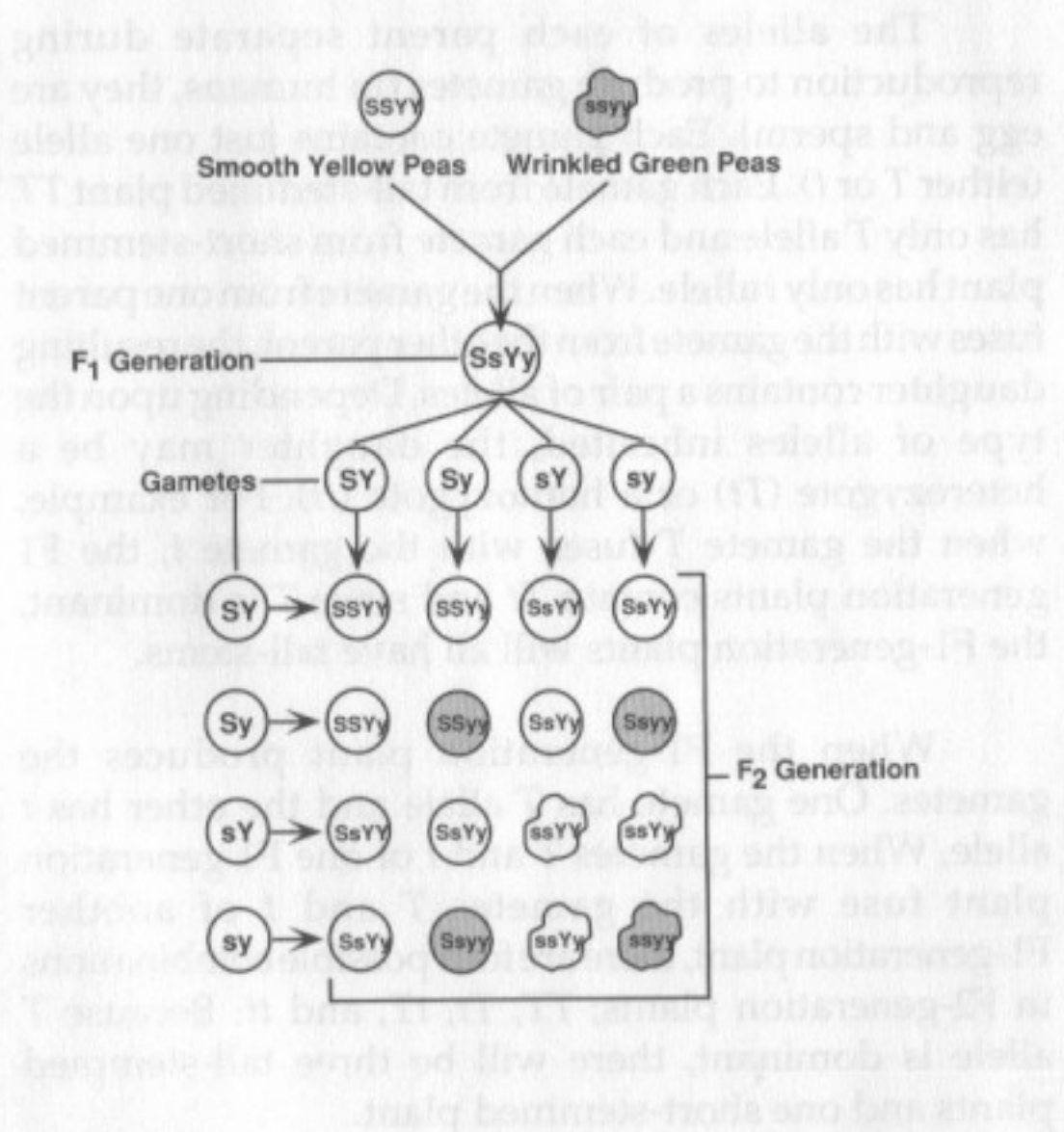


Figure 10.3 Cross fertilization between plants with two characteristic traits.

Comparison of modern genetics to Mendel.

In modern genetics, the **unit factors** that Mendel proposed are **genes**.

Genes and chromosomes occur in pairs (*diploid*=2*n*),

similar to pairs of alleles that he proposed. When meiosis occurs, the daughter cells i.e. gametes, contain just one copy of genes (*haploid*= n).

The dominant alleles are considered **functional** genes. Recessive alleles are genes that are **nonfunctional**. A gene becomes nonfunctional when it cannot express itself. The recessive gene becomes nonfunctional because of a mutation which destroys the information carried by the gene. Functional genes will synthesize a product that gives the organism its characteristic appearance. In the case of the garden pea, the functional gene is the violet color of the flower. A nonfunctional gene is unable to synthesize any product because of mutation and, therefore, is unable to display any characteristic appearance, e.g. white color of the flower.

However, there are others features that Mendel did not notice during his experiments.

For example:

1. The alleles can have **incomplete dominance**. Black color (BB) genes may be incompletely dominant, therefore the heterozygote F_1 -generation (Bb) may show a gray color instead of black (figure 10.4).

2. **Codominance** occurs when both alleles are equally dominant. For

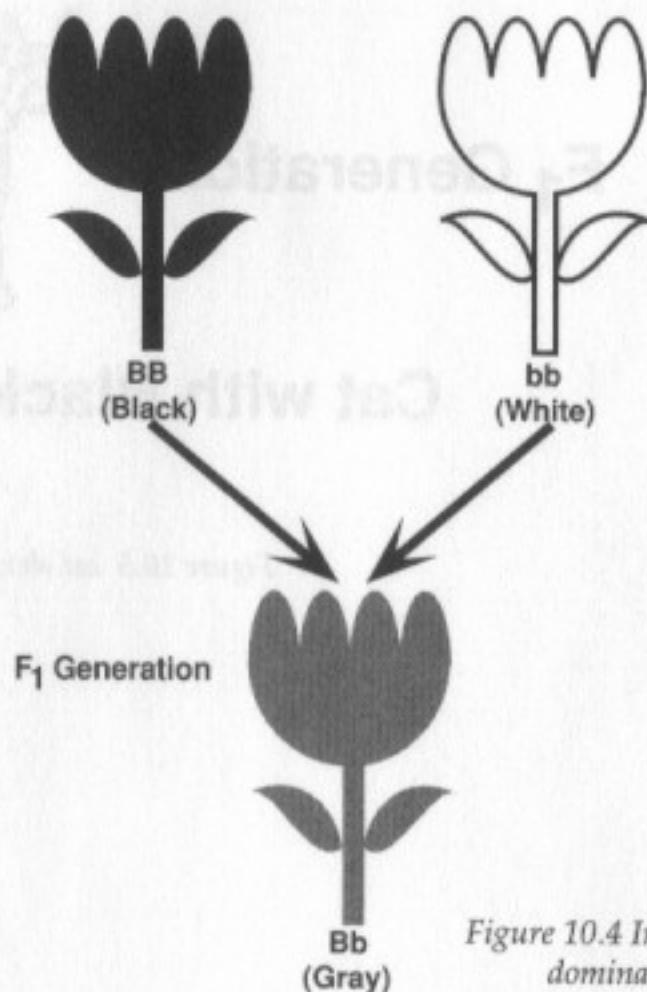


Figure 10.4 Incomplete dominance

example, when a cat with black (BB) fur mates with a cat with white fur (bb). The kitten may show black and white spots instead of having totally black or grey.

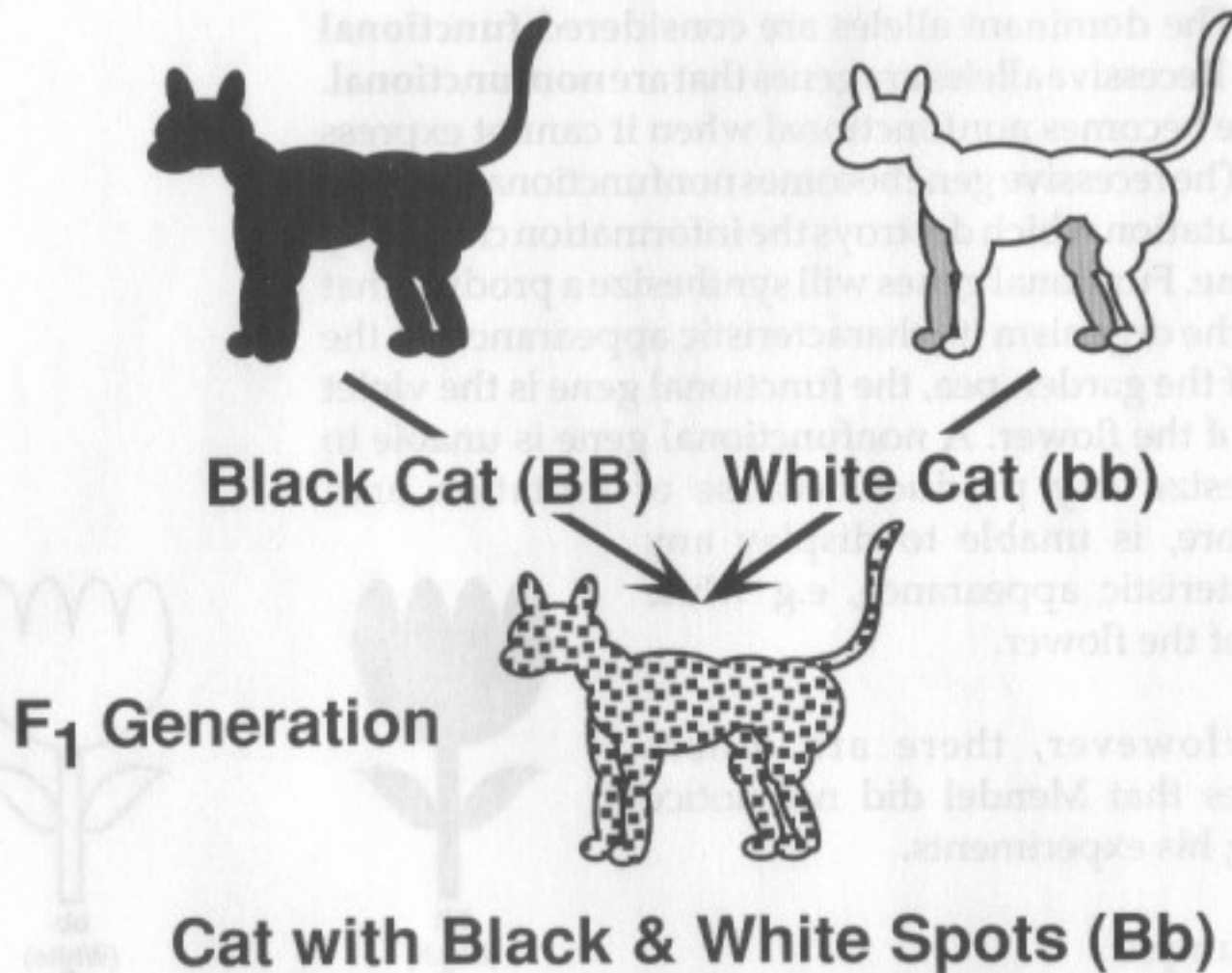


Figure 10.5 cat showing codominant features

Chapter 11

Before a baby is born

This chapter briefly explains how a human baby develops after the union of the father's gamete (sperm) and the mother's gamete (ovum). (Remember, both sperm and ovum are haploid cells).

The sperm and the ovum fuse to form a single cell called a **zygote**. A zygote is a diploid cell. The union of sperm and ovum is called **fertilization**.

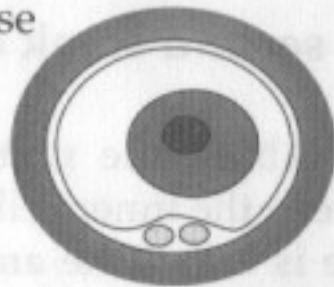


Figure 11.1. Zygote.

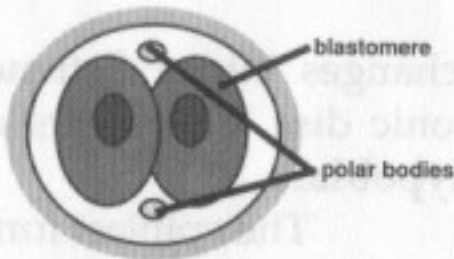


Fig 11.2 Two-cell stage.

The zygote divides by mitosis into two daughter cells. Each daughter cell then divides into four daughter cells and so on. This division by mitosis continues.

About three days after fertilization 12 to 16 cells have formed. The ball of cells resembles a mulberry. It is therefore called a **morula** (morula is a latin name for mulberry).

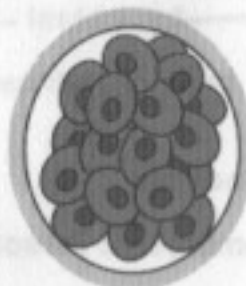


Fig 11.3. Morula.

About the fourth day after fertilization, a fluid-filled cavity is formed in the morula. The cavity is

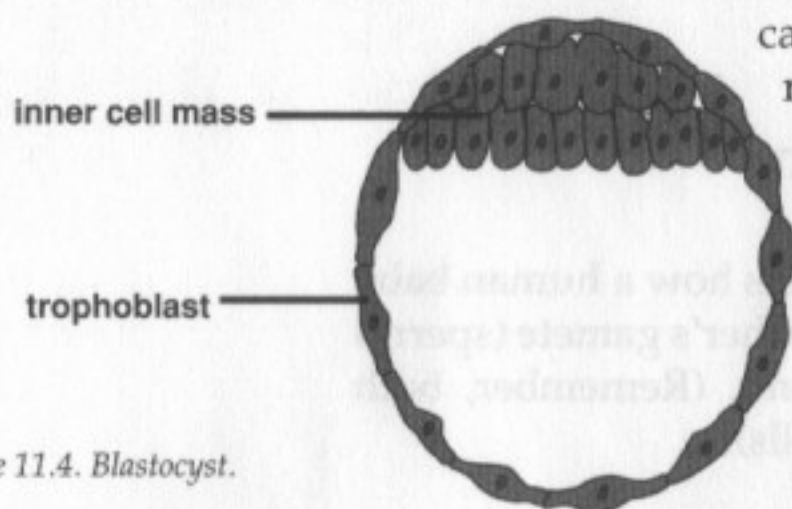


Figure 11.4. Blastocyst.

called a **blastocyst cavity**. The resulting ball of cells with the cavity is called a **blastocyst**. About the sixth to seventh day after fertilization, the blastocyst forms an **inner cell mass** and an outer layer of cells called the **trophoblast**.

The second week of pregnancy

About the ninth day, a small space appears between the inner cell mass and the trophoblast. This space is called the **amniotic cavity** and is filled with amniotic fluid.

The inner cell mass changes into a flattened, two-layered circular embryonic disc. The layers are called the **epiblast** and the **hypoblast**.

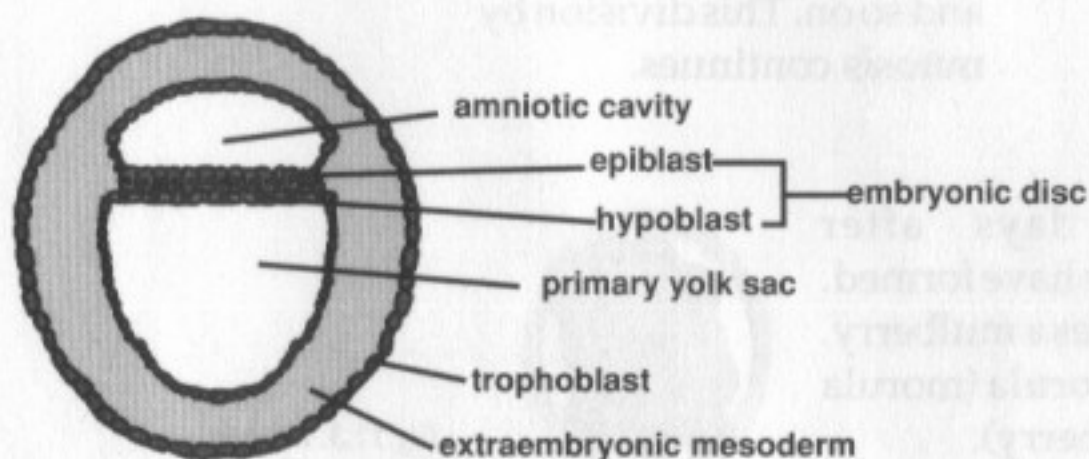


Figure 11.5. Nine-day-old embryo

The epiblast forms the floor of the amniotic cavity. The hypoblast cells grow and form a thin layer of membrane that encloses a cavity called the **primary yolk sac**. The yolk sac is important in the formation of blood.

The thin membrane is called an **exocoelomic membrane**. Some of the cells of the trophoblast develop into a tissue

called the **extraembryonic mesoderm**. The extraembryonic mesoderm develops between the trophoblast cells and the membrane.

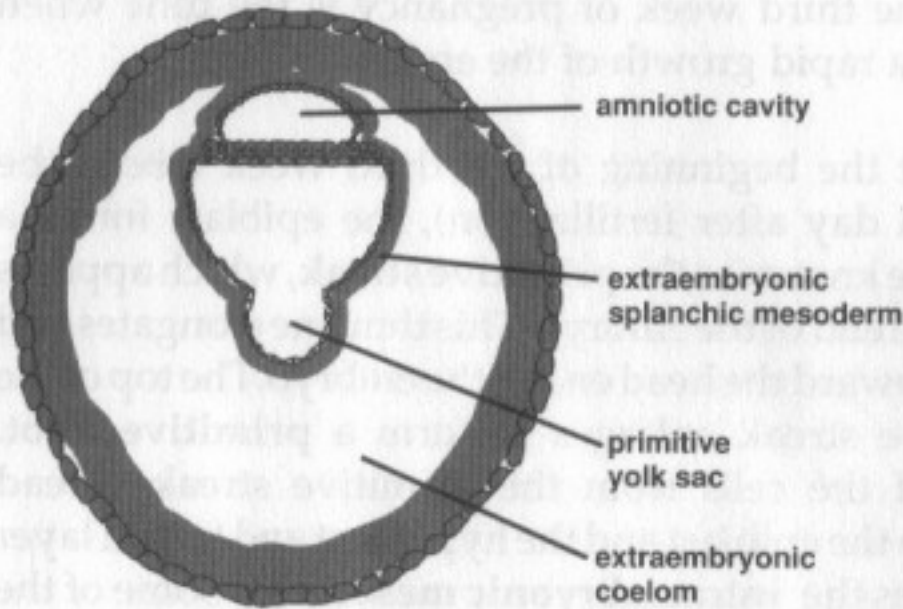


Figure 11.6. Eleven-day-old embryo.

By the eleventh day after fertilization, isolated spaces are formed within the extraembryonic mesoderm. These spaces rapidly fuse to form a large cavity called the **extraembryonic coelom** (coelom means cavity). The extraembryonic coelom splits the extraembryonic mesoderm into two layers. In the meantime, the yolk sac decreases in size.

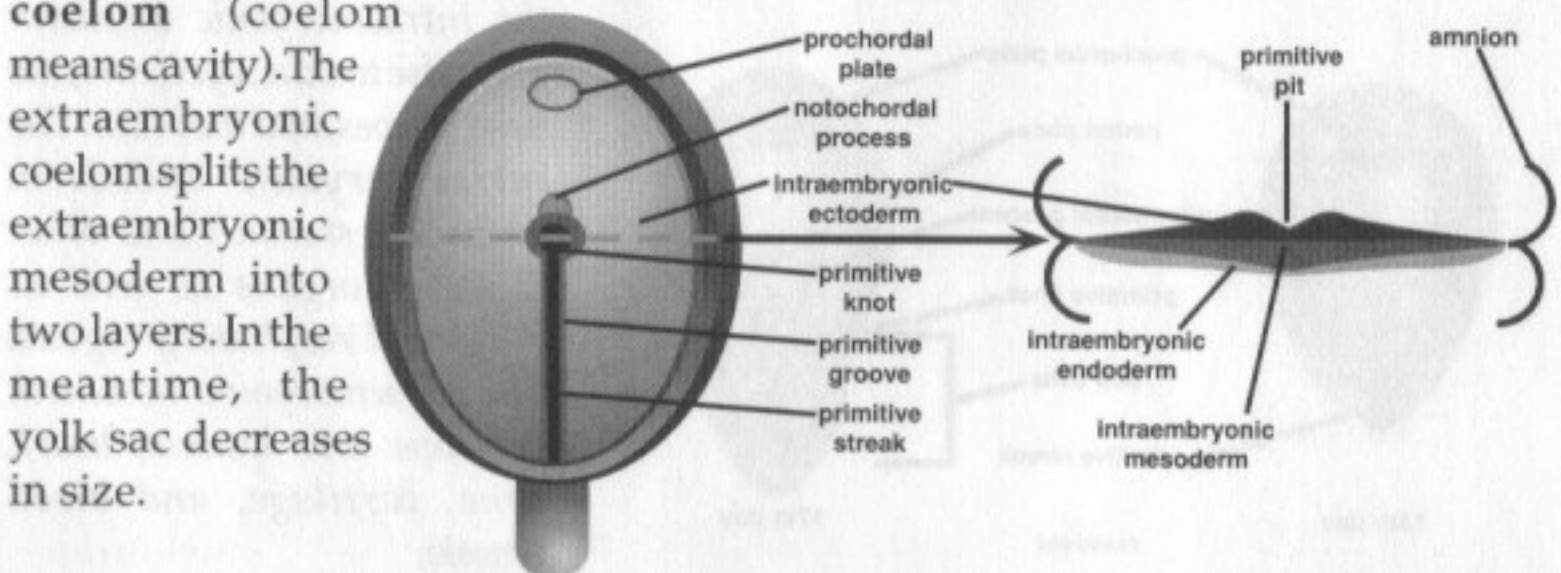


Figure 11.7. Thirteen-day-old embryo.

Three germ layers develop during the third week of pregnancy.

The third week of pregnancy is the time when there is a rapid growth of the embryo.

At the beginning of the third week (about the fifteenth day after fertilization), the epiblast forms a thick line known as the **primitive streak**, which appears at the tail end of the embryo. This thin line elongates and grows toward the head end of the embryo. The top of the primitive streak enlarges to form a **primitive knot**. Some of the cells from the primitive streak spread between the epiblast and the hypoblast and form a layer known as the **intraembryonic mesoderm**. Some of the cells invade and displace the hypoblast cells and form a layer called the **intraembryonic endoderm**. The cells that remain in the epiblast form a layer called the **intraembryonic ectoderm**.

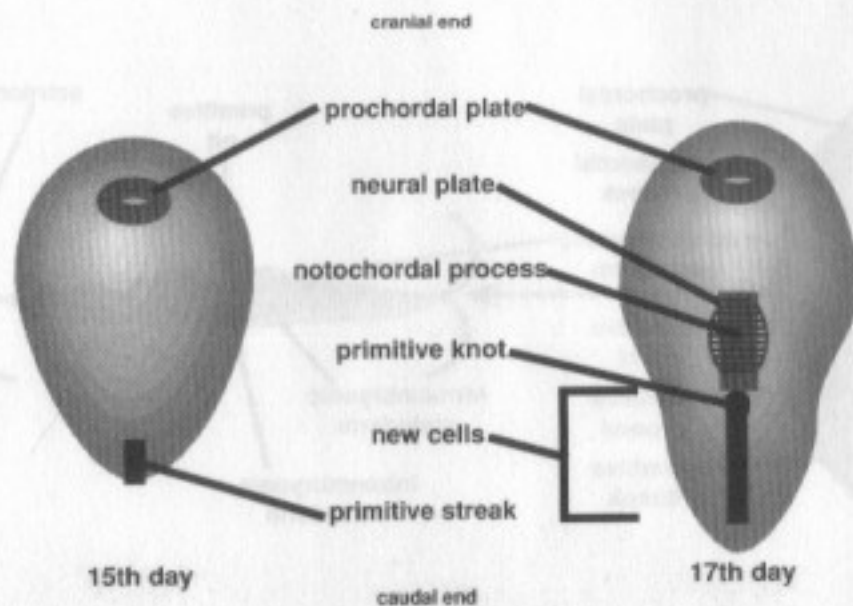


Figure 11.8. Fifteen days after fertilization.

Note: As the embryo gets older the intraembryonic ectoderm gives rise to skin, eyes, ears, the nose, nerves, and the brain. The intraembryonic endoderm forms the stomach and other digestive organs as well as lungs and respiratory organs. The intraembryonic mesoderm develops into muscle, heart, bone, cartilage, and blood vessels.

The cells from the primitive knot continue to grow towards the head end of the embryo. This streak, called the **notochordal process**, grows until it reaches a group of cells called the **prochordal plate**. *The prochordal plate is the future site of the mouth of the embryo.* Just below the primitive streak is a circular area known as the **cloacal membrane**. *The cloacal membrane is the future site of the anus of the embryo.*

The **notochord** is the structure around which the spine bone is formed. The notochord causes the adjacent ectoderm to form a **neural plate**. *The neural plate is the future brain and spinal cord.*

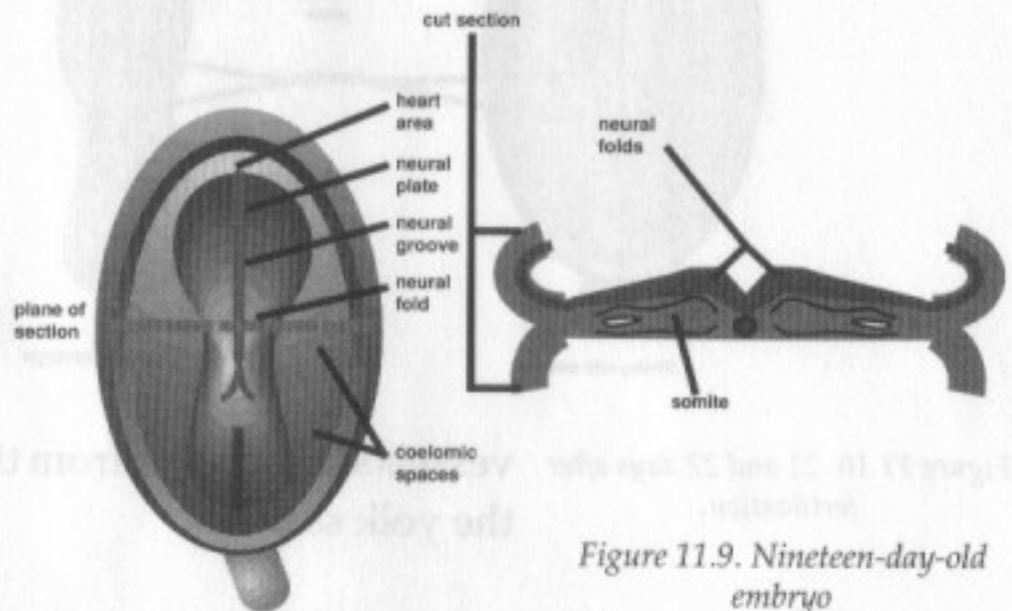


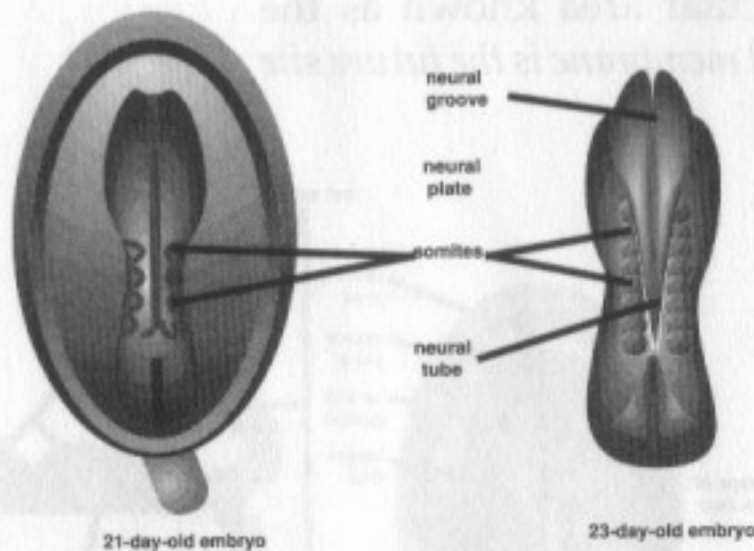
Figure 11.9. Nineteen-day-old embryo

On or about the eighteenth day, the neural plate forms a groove called the **neural groove**, with **neural folds** on each side. Some of the cells form the **neural crest** that later develops into meninges of the brain and spinal cord. Meninges are the linings that surround the brain and spinal cord. The neural crest also develops into some of the skull bones and adrenal glands.

By the end of third week of pregnancy the notochord is almost completely formed. The neural folds fuse to form a tube which later closes.

At the end of the third week, as the neural tube is formed, the mesoderm beside the tube forms into paired cuboidal bodies called **somites**. Eventually 42 to 44 pair of somites develop.

The somites later develop into spine, ribs, sternum, and skull bones and associated muscles.



At the beginning of the third week blood and blood

Figure 11.10. 21 and 22 days after fertilization.

vessels start to form from the mesoderm that surrounds the yolk sac.

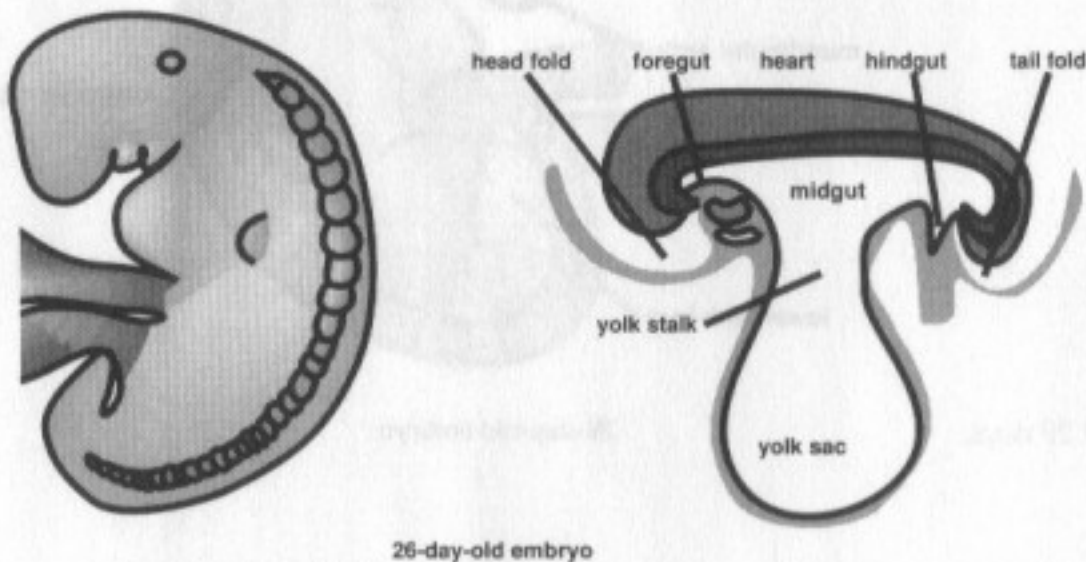
Cells from the area near the head of the embryo develop into paired heart tubes before the end of the third week and fuse to form a primitive heart. By the end of the third week the heart begins to beat.

About the ninth week, the intraembryonic coelom is divided into three body cavities (1) the pericardial cavity, which surrounds the heart, (2) the pleural cavities which surround the lungs, and (3) the peritoneal cavity which contains the stomach and the bowels in the abdomen.

The fourth to eighth week old embryo

The organs of the body are formed during the fourth to eighth week. By the end of the eighth week, the embryo has taken on a remarkably human appearance.

The three-layered embryonic disc elongates and folds into a head fold and a tail fold. The embryo also folds laterally on both sides. During folding of the head region, part of the yolk sac becomes the **foregut** which later develops into the larynx and the esophagus. During folding of the tail region part of the yolk sac becomes the **hindgut** which later becomes the large intestines and rectum. Part of the yolk sac becomes the **midgut**, which later develops into the stomach and small intestines. The three layers, the ectoderm, mesoderm and endoderm develop into all the structures and organs mentioned earlier.



26-day-old embryo

Figure 11.11. Embryo at 4 weeks.

About the twenty fourth day, a **mandibular arch** develops. This arch later develops into a lower jaw.

Thickenings called **otic pits** appear in the head region. Another thickening, the **lens placode**, also appears. Otic pits later develop into inner ears.

About the twenty-sixth day upper limb buds appear. The upper limb buds develop into arms. About day thirty-two the limb buds appear paddle-shaped. Lower limb buds also appear. Lower limb buds later develop into legs. The tail end gradually shortens.

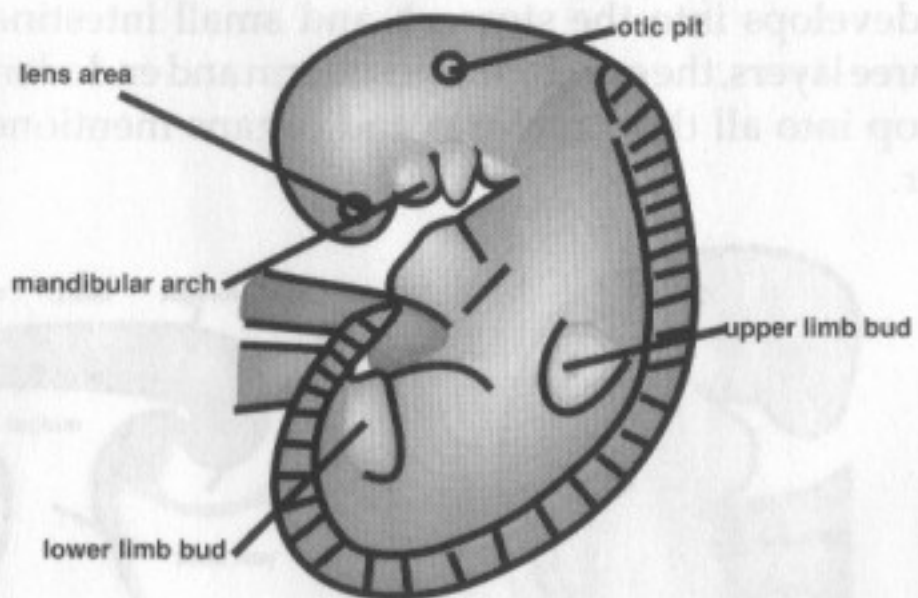


Figure 11.12. Embryo at 29 days.

29-day-old embryo

About day forty-one fingers starts to develop in the upper limbs. About the forty-eighth day toes also start to develop in the lower limbs (figure 11.13, pg 87).

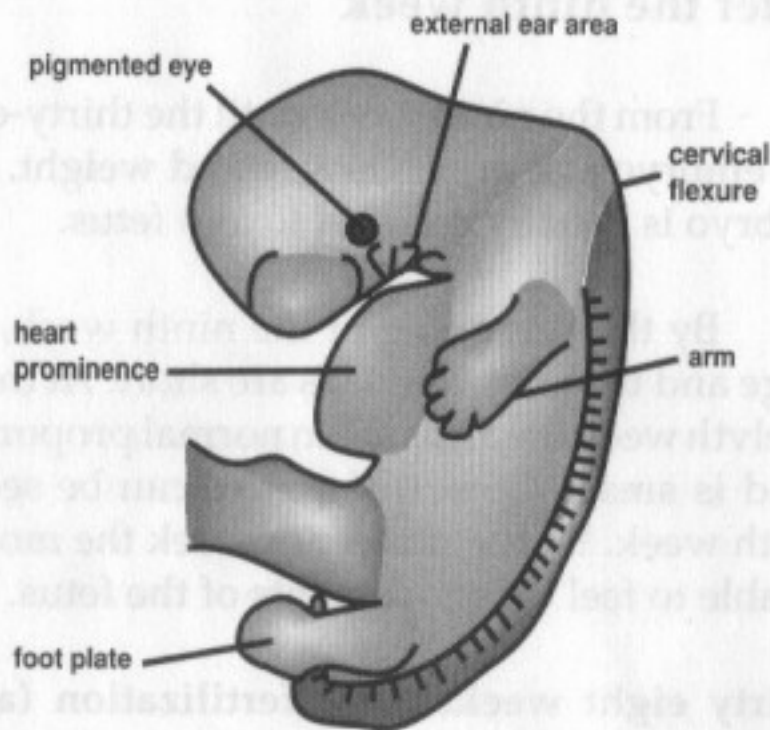


Figure 11.13. 41 days.

About day fifty-one eyelids are seen in the otic pits and external ears begin to develop.

About day fifty-six the head is round, but large. The eyes are open. External ears are now taking their final shape. Sex appearances start to show but are not distinct.

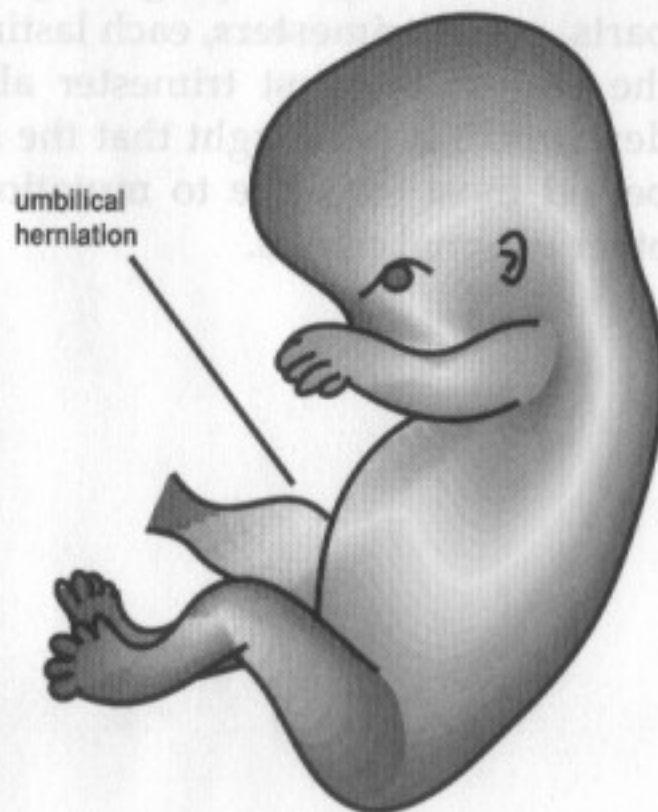


Figure 11.14. Embryo at day 56.

After the ninth week

From the ninth week until the thirty-eighth week the embryo just grows in size and weight. By now the embryo is usually referred to as a fetus.

By the beginning of the ninth week, the head is large and the arms and legs are short. At the end of the twelfth week the arms are in normal proportion and the head is smaller. Sex differences can be seen after the ninth week. By the nineteenth week the mother should be able to feel the movements of the fetus.

Thirty eight weeks after fertilization (and usually forty weeks after the last menstruation) the baby is born.

Commonly, the pregnancy is divided into three parts, called **trimesters**, each lasting three months. By the end of the first trimester all major organs are developed. It is thought that the first trimester is the period most sensitive to mutations from drugs and other external causes.



References

1. The DNA Story (1981)
Watson, J.D., and Tooze, J.
Freeman Publishing
2. Principles of Genetics (1991)
Gardner, E.J., Simmons, M.J., and Snustad, D.P.
John Wiley & Sons, Inc.
3. Recombinant DNA (1992)
Watson, J.D., Gilman, M.
W.H. Freeman and Company
4. Genetics - A Molecular Approach (1990)
Brown, T.A.
Chapman and Hall
5. Molecular Cell Biology (1990)
Danell, J., Lodish, H., Baltimore, D.
W.H. Freeman and Company
6. Molecular Biology of the Cell (1989)
Alberts, B., Bray, D., Lewis, J., Raff, M.,
Roberts, K., Watson, J.D.
Garland Publishing, Inc.