2023/2024

INTRODUCTION TO BIOTECHNOLOGY

2nd Year

Molecular Biotechnology

Prepared by:

Dr. Amr Mohamed Ali

Lecturer of Genetics and Molecular Biology

Zoology Department, Faculty of Science

PREFACE

Biotechnology can be considered as the "automobile" of the 21st century. It is affecting almost every aspect of society in the same way as the first mass production automobile changed the world in the late 1800s. Many historians view that automobile as a phenomenal technology that brought about unparalleled global prosperity. Biotechnology is likely to bring global prosperity by providing more effective ways to grow foods, manufacture commercial products, produce energy, and treat diseases. The number of new biotechnology applications that make their way into society is increasing rapidly every year. More and more government and university laboratories are dedicating resources to biotechnology research and development. Biotechnology is becoming an increasingly popular career choice for college students enrolled in biology, chemistry, engineering, and physics programs. Many law schools offer courses and specialties in biotechnology-related areas. Allied health professionals must now receive continuing education training to understand the growing number of medical biotechnology applications they are encountering today and in the near future.

There have been considerable benefits and risks to every technology that has been introduced throughout the world in the past three centuries. For example, the automobile paved the way for rapid transportation that spurred the growth of suburbs and fast-food restaurants. However, the automobile is blamed for depleted fossil fuel reserves and for considerable amounts of air pollution. The benefits of current biotechnology applications include improvements in agricultural products, safer medicines, precise treatments for genetic disorders, accurate medical diagnosis technologies, environmentally cleaner ways of producing commercial chemicals and crops, and alternatives to fossil fuels.

INTRODUCTION

Biotechnology is the youngest of the sciences and is increasing in knowledge at an unprecedented rate. It is the fastest growing technical discipline and has probably gained more information per year than any other field of science. Advances in biotechnology even outpace new developments in computer science. Because of the rapid advance, biotechnology is called a revolutionary science that outpaces that ability for people to keep up with an understanding of applications in society. The term biotechnology was first used by Hungarian engineer Karoly Ereky in ´ 1919. His use of the term varies somewhat from its meaning today. Ereky used biotechnology to describe the industrial production of pigs by feeding them sugar beets as an inexpensive large-scale source of nutrients. He then generalized the term to all areas of industry in which commercial products are created from raw materials with the aid of organisms. Ereky predicted a biochemical age that rivaled the societal impacts of the Stone and Iron Ages.

The science of biotechnology is an amalgamation of biology, chemistry, computer science, physics, and mathematics. Many scientists who work in biotechnology fields have a diversity of skills that bring together two or more science disciplines. Biotechnology is also practiced as a working relationship between two or more scientists who collaborate on projects by sharing their expertise and experiences. Certain types of biotechnology involve many specialized techniques which only a few people can perform. Yet, other procedures and scientific instruments used in biotechnology are fairly simple. The biotechnology concepts and techniques taught only to graduate and postdoctoral students in the 1970s are now covered in high school science classes.

Biotechnology Definition:

"The study of living tools"- is used in agriculture, food processing, industrial production, environmental cleanup and medicine. A set of modern tools that utilize living organisms or parts of it cell or tissue or genes/DNA to make or modify or improve plants or animals or develop microorganisms for specific use or their large scale production. "Utilization of organisms or its organells or biological process to make product or to solve problems for the welfare of mankind." The Convention on Biological Diversity (CBD, 2000) biotechnology means "any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use".

History of Biotechnology:

Biotechnology seems to be leading a sudden new biological revolution. Biotechnology is NOT new. Man has been manipulating living things to solve problems and improve his way of life for millennia. The term biotechnology was coined in 1919 by Karl Ereky, an Hungarian engineer. At that time, the term meant all the lines of work by which products are produced from raw materials with the aid of living organisms. Although now most often associated with the development of drugs, historically biotechnology has been principally associated with food such as malnutrition and famine. The history of biotechnology begins on brewing techniques for beer. Early agriculture concentrated on producing food. Plants and animals were selectively bred, and microorganisms were used to make food items such as beverages, cheese and bread. The ancient Egyptians made wine using fermentation techniques based on an understanding of the microbiological processes that occur in the absence of oxygen. Egyptians also applied fermentation technologies to make dough rise during bread making. The late eighteenth century and the beginning of

the nineteenth century saw the advent of vaccinations, crop rotation involving leguminous crops, and animal drawn machinery. The end of the nineteenth century was a milestone of biology. Microorganisms were discovered, Mendel's work on genetics was accomplished, and institutes for investigating fermentation and other microbial processes were established by Koch, Pasteur, and Lister.

Biotechnology at the beginning of the twentieth century began to bring industry and agriculture together. In 1928, Alexander Fleming discovered the mold *Penicillium*. In 1940, penicillin became available for medicinal use to treat bacterial infections in humans. The biotechnical focus moved to pharmaceuticals. In 1953, James Watson and Francis Crick's were discovered the structure of DNA. The field of modern biotechnology is generally thought of as having been born in 1971 when Paul Berg's experiments in gene splicing had early success. Herbert W. Boyer and Stanley N. Cohen significantly advanced the new technology in 1972 by transferring genetic material into a bacterium, such that the imported material would be reproduced. In 1978, Boyer was able to take pieces of human DNA and isolate a gene for insulin using biotechnology. In the 1980s, testing of biotechnology-derived foods began, and after its FDA approval in 1994, the FlavrSavr® tomato gave consumers a more flavorful tomato that stays fresh longer. Today's biotechnology has its "roots" in chemistry, physics, and biology.

Traditional and New Biotechnology: Although the term biotechnology is of recent origin, the discipline itself is very old. Man began employing microorganisms as early as 5000 BC for making wine, vinegar, bread etc. Some aspects of biotechnology are as ancient and familiar as adding yeast to bread dough; others are as recent and unfamiliar as genetic engineering.

Traditional: Biotechnology has been employed by humans for millennia. The ability of microorganisms to produce acids and gasses as a result of normal cell metabolism has been taken advantage of to make new and exciting foods for generations. Examples include production of beer, cheese and bread.

New Biotechnology: Recent developments in molecular biology have given Biotechnology a new meaning, new horizon and new potential through use of recombinant DNA technology. New biotechnology to modify the genetic material of living cells to produce new substances or perform new functions. Gene technology or genetic engineering allows the biologist to take a gene from one cell and insert it into another cell which may be plant, animal or microbial (bacterial or fungal), or to produce new combinations of genes.

Scope & Importance of Biotechnology:

Biotechnology has rapidly emerged as an area of activity having a marked realized as well as potential impact on virtually all domains of human welfare, ranging from food processing, protecting the environment, to human health. As a result, it now plays a very important role in employment, production and productivity, trade, economics and economy, human health and the quality of human life throughout the world. This is clearly reflected in the creating of numerous biotechnology companies throughout the world, and the movement of noted scientists, including Nobel laureates, to some of these companies. The total volume of trade in biotechnology products in increasing sharply every year, and it would soon become the major contributor to the world trade. Many commentators are confident that the 2lst century will be the century of biotechnology, just as the 20th century was the era of electronics.

THE GENETIC MATERIALS

CHROMOSOMES

History: W. Hofmeister in 1848, discovered nuclear filaments in the nuclei of pollen mother cells of *Tradescantia*. First accurate count of chromosomes was made by W. Flemming in 1882, in the nucleus of a cell. In 1884, W. Flemming, Evan Beneden and E. Strasburger demonstrated that the chromosomes double in number by longitudinal division during mitosis. Beneden in 1887 found that the number of chromosomes for each species was constant. The term "Chromosomes" was coined in 1888 by W. Waldeyer for the nuclear filaments. W.S. Sutton and T. Boveri suggested role of chromosomes in heredity in 1902, confirmed by Morgan in 1933.

- **The structure of chromosomes varies in viruses, prokaryotes and eukaryotes :**

A. **Viral chromosome-** In viruses there is a single chromosome bearing a single nucleic acid molecule (DNA or RNA) surrounded by a protein coat called Capsid. It may be linear or circular. The viruses having DNA as genetic material are called DNAviruses and those having RNA as genetic material are known as RNA viruses. A limited amount of genetic information is present in the viral chromosome which codes for little more than the production of more virus particles of the same kind in the host cell. In RNA viruses, often the RNA directs the synthesis of DNA complementary to itself by reverse transcription in the host. The RNA is then transcribed by the DNAfor the formation of new virus particles. Such ribovirus is called retrovirus. TheAIDS causing virus is a retrovirus.

B. **Prokaryotic chromosomes-** Prokaryotic chromosome (e.g., bacteria) has asingle and circular two-stranded DNA molecule which is not enveloped by any membrane. It lacks proteins and is in direct contact with the cytoplasm. The

bacterial chromosome is packed into the nucleoid by some RNA that appears to form a core. It is attached to plasma membrane permanently at least at one point. In addition to the main chromosome some extra-chromosomal DNA molecules may also be present in most of the bacterial cells they are also double stranded and circular but are much smaller in size. They are known as plasmids. The plasmid may occur independently in the cytoplasm of calls or may also be found in association of main chromosomal DNA and called as episome.

C. **Eukaryotic chromosomes-** The eukaryotic chromosomes are present in nucleus and in certain other organelles, like mitochondria and plastids. These chromosomes are called nuclear and extra nuclear chromosomes respectively. Nuclear chromosomes are double stranded long DNA molecules of linear form. Proteins are associated with them. They are surrounded by nuclear envelope. More DNA is involved in coding far more proteins than the prokaryotic chromosomes. Extra nuclear chromosomes are present in mitochondria and plastids. They are double stranded short DNA molecules of circular form. They lack protein association. Less genetic information is available for the synthesis of only some particles of proteins for the organelles containing them. Other proteins are received from the cytoplasm where they are synthesized under the direction of nuclear chromosomes.

➢ **Material of Chromosomes**

The chromatin material of the eukaryotic chromosomes according to its percentage of DNA, RNA and proteins and consequently due to its, staining property has been classified into following by classical cytologists:

a) Euchromatin

The euchromatin is the extended form of chromatin and it forms the major, portion of chromosomes. The euchromatin has special affinity for basic stains and is genetically active because its component DNA molecule synthesizes RNA molecules only in the extended formof chromatin.

b) Heterochromatin

The heterochromatin is a condensed intercoiled state of chromatin, containing two to three times more DNA than euchromatin. However, it is genetically inert as it does not direct synthesize RNA (i.e., transcription) and protein and is often replicated at a different time fromthe rest of the DNA.

Recent molecular biological studies have identified three kinds of heterochromatins, namely constitutive, facultative, and condensed heterochromatin. The constitutive heterochromatin is present at all times and in the nuclei of virtually all the cells of an organism. In an interphase nucleus, it tends to clump together to form chromocenter or false nucleoli. In Drosophila, for example, most pupal, larval and adult cells contain large blocks of constitutive heterochromatin that lie adjacent to centromeres. Constitutive heterochromatins contain highly repetitive satellite DNA which is late replicating, it fails to replicate until late in the 5-phase and is then replicated during a brief period just before the G2. The facultative heterochromatin reflects the existence of a regulatory device designed to adjust the "dosages" of certain genes in the nucleus

➢ Morphology of Chromosomes

During the interphase stage, the eukaryotic chromosomes are extended into long and thin chromatin fibers where they lie criss-cross to form the **chromatin reticulum**. They replicate in the S-phase and become double. At this stage they consist of two chromatids that are held together at one point called **centromere.** At the time of cell

division, the chromosomes condense and tightly coil up and become distinct at metaphase stage. The eukaryotic chromosomes vary in number, size, shape and position but they have remarkably uniform structure.

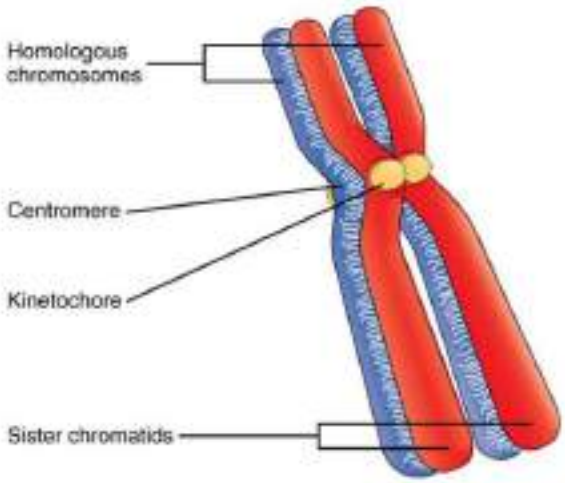


Figure1: A Homologous Pair of Chromosomes with their Attached Sister Chromatids.

- **Number of Chromosomes**

Eukaryotic chromosomes vary in number from two to a few hundred in different species. In a species all the individuals have same number of chromosomesin all of their cells, except the gametes. Since the chromosome number is constant for a species, it is helpful in determining and taxonomic position of thespecies.

- **Size of Chromosomes**

In a **species all the chromosomes are not of the same size.** Their size also varies from species to species. The particular chromosome of a species however has more or less a constant size. The organisms having fewer chromosomes have large sized chromosomes than those having many. Generally, plant chromosomes are larger than animal chromosomes and among plants the monocots have larger chromosomes than the dicots.

➢ Shape of Chromosomes

The chromosomes at metaphase stage look like slender rods that may be straight or curved to form an arc or a letter S. In anaphase stage they may assume J or V shapes, depending upon the position of the centromere.

➢ Position and chemical composition of Chromosomes

In a nucleus each chromosome is independent of all the other chromosomes in its location. Thus, they may occupy any region of the nucleus. The chromatin in the eukaryotic chromosome consists chemically of about 35% DNA, about 60% proteins, about 5% RNA, some metal ions and certain enzymes.

➢ Structure of Chromosomes

At **metaphase stage**, since the chromosome is a highly condensed nucleoprotein filament, it contains two greatly coiled sister chromatids. These chromatids that lie side by side along their length, are held together at a point called centromere, an area of the narrow region also called **primary constriction** of the metaphase chromosome. At the centromere each chromatid has a darkly staining, disc like, fibrous structure, called **kinetochore**, to which spindle microtubules attach during cell division. Kinetochores are the sites where force is exerted to pull the chromatids towards the poles. One or more chromosomes may have additional narrow regions called the **secondary constrictions**. The part of the chromosome separated by secondary constrictions is termed as **satellite**. A chromosome with a satellite is called **sat chromosome**. The size and the shape of the satellite remain constant for a species. Secondary constrictions are associated with the nucleoli and are known as the **nucleolar organizers**. The chromosomes which have nucleolar organizing regions areknown as the **nucleolar chromosomes**. **Ends-** The ends of chromosomes are called **telomeres**. The function of telomere varies from the rest of the

chromosome. On exposure to X-rays a chromosome may break and its pieces may rejoin, but no segment connects to the telomere, showing that the telomere has a polarity, and it, somehow "seals" the end.

➢ Ultra-structure of Chromosomes

A chromatid contains a very fine filament called chromonema whichis a single, long, double stranded DNA molecule. It is wrapped around histones to form **nucleosomes**. The nucleosome and non-histone proteins together form the chromatin fiber. The chromatin fiber has reactive groups, probably H1 histone molecules, which act as "folders" and crosslink the chromatin fiber changing it into a great coiled, compact metaphase chromatid.

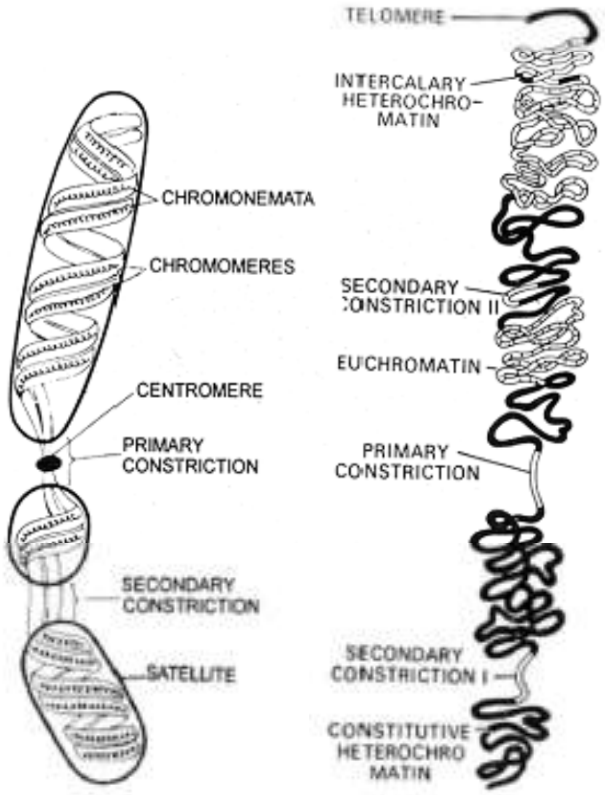


Figure 2: Detailed schematic structure of chromosomes.

Types of Chromosomes

On the basis of the position and number of centromeres, chromosomes are classified:

(i) **Metacentric:** In metacentric chromosomes the centromere is at the middle of the chromosome, and the arms are equal. In anaphase the chromosome appears V-shaped. For example: human chromosome no. 3

(ii) **Submetacentric:** In such chromosome, the centromere is near the center of the chromosome, and the arms are slightly unequal and in anaphase the chromosome appears J or L shaped. For example: Human chromosome No. 1.

(iii) **Acrocentric:** In this type the centromere is near one end of the chromosome, and the arms are very unequal. For example: Human chromosome No. 4 & 5.

(iv) **Telocentric:** The centromere is at one end in such chromosomes, and the arms are on one side only. The chromosome remains rod shaped in anaphase also

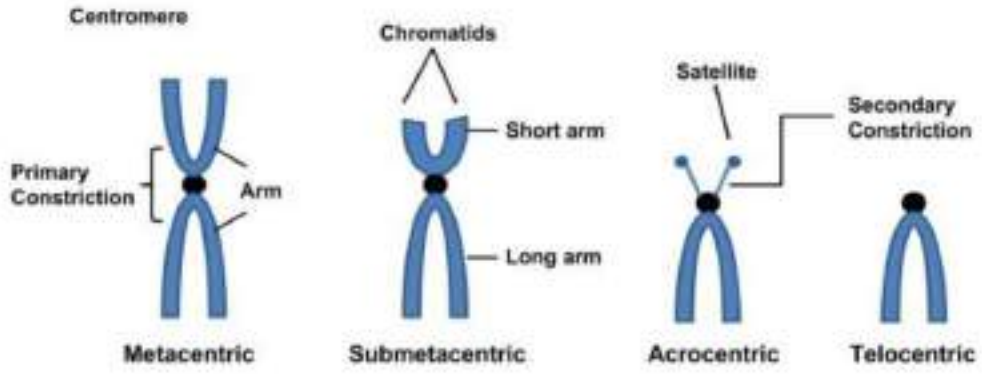


Figure 3: Types of chromosomes based on the position and number of centromeres.

Depending upon the number of centromeres there are three types of chromosomes:

(i) **Acentric:** The chromosome is without a centromere, which is formed by breakage of the chromosome. It does not attach to spindle microtubules, so it is lost in the cell division.

(ii) **Monocentric:** It is the chromosome with a single centromere, and it is the most common type.

(i) **Dicentric:** It is the chromosome with two centromeres and is formed by the fusion of two chromosome segments each having a centromere. It is unstable and may break when the two centromeres are pulled to opposite poles in mitosis.

➢ **Functions of Chromosomes**

1- Chromosomes carry hereditary characters from parents to offspring.
2- They direct the synthesis of structural proteins and thus, help cell grow, and divide.
3- By directing the formation of necessary enzymes, they control metabolism.
4- They guide cell differentiation during development.
5- They form nucleoli at nucleolar organizer sites in daughter cells.
6- They produce variations through changes in their genes and contribute to the evolution of the organisms.
7- They play role in sex determination.
8- They maintain the continuity of life by replication.

➢ **Special types of chromosomes**

Some tissues of certain organisms contain chromosomes, which differ significantly from normal chromosomes in terms of either morphology or function. Such chromosomes are referred to as special chromosomes.

- **Giant Chromosomes**

Giant chromosomes are special, enormously enlarged chromosomes about 100 times thicker than the ordinary mitotic chromosomes. These are seen in certain tissues of varied groups of animals and plants. They are easily visible under light microscope. The giant chromosomes are of two types: polytene and lampbrush.

(A) Polytene Chromosomes

Polytene chromosomes were first observed by Balbiani (1881) in Chironomus (a dipteran larva). Because of their large size showing numerous strands these are named as polytene chromosomes by Kollar. These banded chromosomes occur in the larval salivary glands (salivary gland chromosomes), midgut epithelium, and rectum and Malpighian tubules of various genera of dipterans.

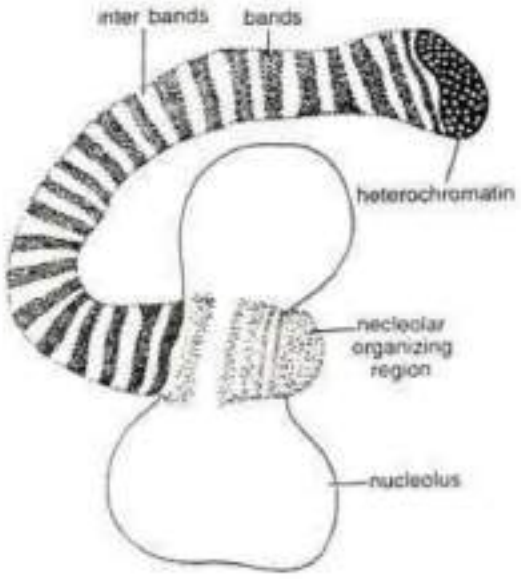


Figure 4: Structure of polytene chromosome showing nucleolar part.

These chromosomes are about 100-200 times larger than those of somatic chromosomes. They are roughly cylindrical and exhibit a distinct pattern of transverse striated structures consisting of alternate darkly staining band and light staining interbands. Dark bands are rich in DNA along with a small amount of RNA and basic proteins. They are genetically active. The inter-bands contain less of DNA but more acidic proteins and hence they are less active. The polytene chromosomes are formed by repeated replication of DNA without division of chromosome into daughter chromosomes. This amplification without separation is called

polytenization. As a result, a thick bundle of parallel DNA molecules all having the same banding pattern across them is produced. Thus, there can be as many as several thousands of chromonemata in a giant chromosome.

Functions of the Giant Polytene Chromosomes

- Polytene chromosomes carry genes which ultimately control physiology of an organism. These genes are formed of DNA molecules.
- These chromosomes also help in protein synthesis indirectly. The RNA present in the nucleolus serves as a means of transmission of genetic information to the cytoplasm, leading to the formation of specific protein.

(B) Lampbrush Chromosomes

These are the largest chromosomes which can be seen with naked eyes and are found in yolk rich oocytic nuclei of certain vertebrates such as fishes, amphibians, reptiles, andbirds. They are characterized by the fine lateral loops, arising from the chromomeres, during first prophase of meiosis. Because of these loops they appear like brush; that is why they are called lampbrush chromosomes first discovered by Flemming in 1882 and described in shark oocytes by Ruckert (1892).

Lampbrush chromosome consists of longitudinal axis formed by a single DNA molecule along which hundreds of beads like chromomeres are distributed. Two symmetrical lateral loops (one for each chromatid) emerge from each chromomere, which are able to expand or contract in response to various environmental conditions. About 5 to 10% of the DNA is in the lateral loops. The axis having compacted DNA and tightly associated proteins is transcriptionally inactive. The loops consist of uncompacted DNA and proteins but have a good amount of RNA and they are transcriptionally active. A chromomere and its associated loop correspond with one gene. In lampbrush chromosomes the DNA loops are the sites

of intensive RNA synthesis. rRNA and mRNA are synthesized in large amount and the transcription of rRNA causes the enlargement of nucleolus, or formation of numerous additional nucleoli. Due to the synthesis of large amounts of proteins, fats, carbohydrates, and other molecules in the cytoplasm needed for further development of the embryo, the oocyte grows in size. Synthesis of proteins occurs near the loops.

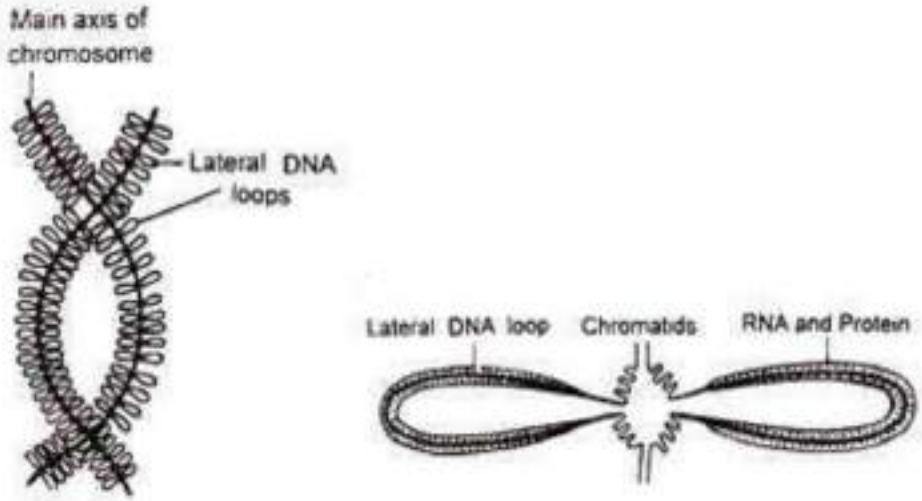


Figure 5: Detailed structure of lampbrush chromosome.

Functions of Lampbrush Chromosome:

- Involved in the synthesis of RNA and proteins by their loops.
- Probably help in the formation of certain amount of yolkmaterial for the egg.

(C) Iso-chromosomes:

A chromosome with two identical arms and identical genes is called as isochromosome. The arms are mirror images of each other. IT is thought to arise when a centromere divides in the wrong plane yielding two daughter chromosomes, each ofwhich carries the information of one arm only but present twice. At meiosis isochromosomes pair in three different ways. (i) Internal pairing (ii) Fraternal pairing (iii) Normal pairing

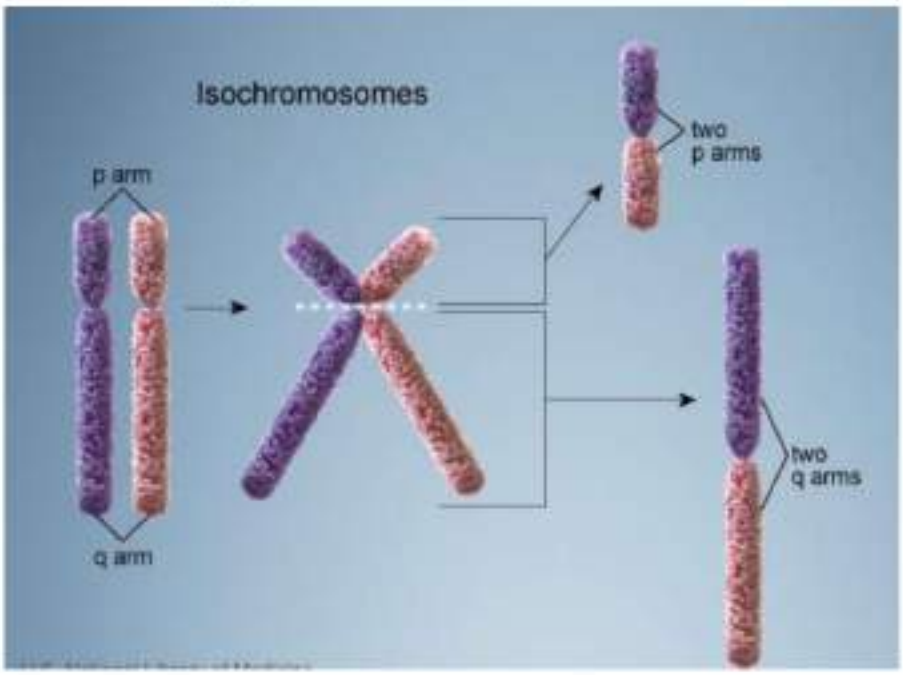


Figure 6: Structure of isochromosomes chromosome.

In internal pairing, the two arms of the isochromosomes pair with each other. In fraternal pairing, one or both of the arms of the isochromosomes pair with a homologous arm of another chromosome. In normal pairing, the isochromosome pairs with another one just like it.

(D) 'B' chromosomes:

It is a particular kind of supernumerary chromosome that may or may not be found in organisms as extra chromosomes over and above the standard diploid or polyploidy chromosome complements. The standard complements are called 'A' chromosome. The 'B' chromosomes found in natural population are recognized on the basis of following characteristics.

- They are dispensable (not found is all the individuals of the species or all the cells of the organisms)
- They are not homologous with any of the basic 'A' chromosomes.

- Their inheritance is non Mendalian.
- They are usually smaller than the 'A' chromosomes.
- Generally, they are genetically inert.
- When it presents in higher number, they suppress the vigour and fertility.
- Their origin and functions are largely unknown.

The most significant effect of 'B' chromosome is on seed and pollen fertility. Flowering time is generally delayed by 'B' chromosomes and has negative on sequences for the organism as they have deleterious effect because of abnormal crossing over during meiosis.

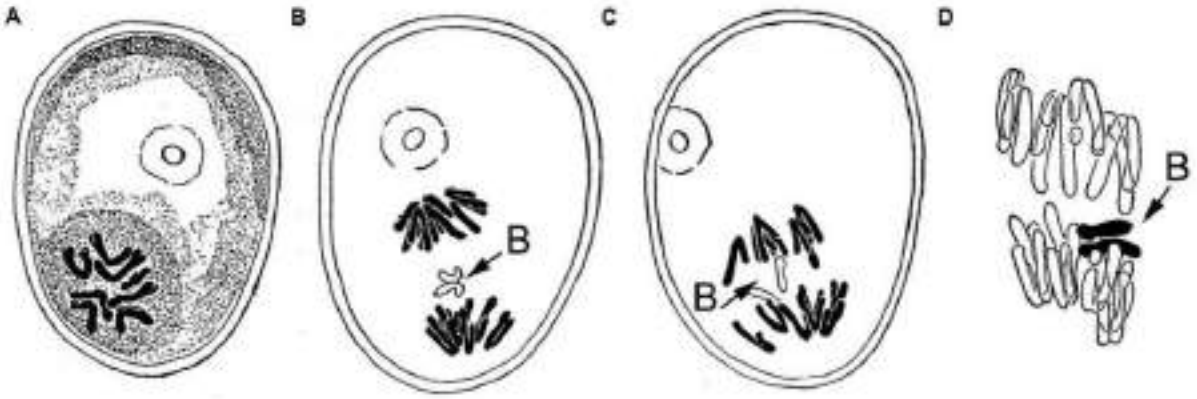


Figure 7: (A) metaphase, (B) lagging B chromosome due to non-disjunction. (C) Disjoined sister chromatids of the B chromosome going to different poles, (D) chromosome drive occurs, the future generative nucleus receives both sister chromatids of the B

(E) Ring chromosome:

The chromosomes of higher organisms usually have two ends and do not form a continuous ring. However, the chromosomes of lower organisms such as prokaryotes. (*E.,coli*) normally have ring shaped chromosomes. Often such chromosomes are referred to as gonophores. Which are more than 1 mm in length and consists of a single DNA molecule.

Chromosomes in higher organisms are not naturally ring shaped. However, ring chromosomes have been detected in humans, *Drosphila* and certain plant species. Ring chromosomes were most thoroughly studied in maize by Mc Clintock. Normal chromosomes do not form rings because they are believed to have telomeres on each end. Telomeres prevent the union of chromosome arms into ring formation. A chromosome can form a ring chromosome by fusion of the raw ends onlyif it has two terminal deletions producing centric segment with two raw ends and two acentric fragments. A ring chromosome lacks the genetic information that was carried by the terminally deleted fragments. Ring chromosomes are meiotically unstable, and they are associated with several syndromes.

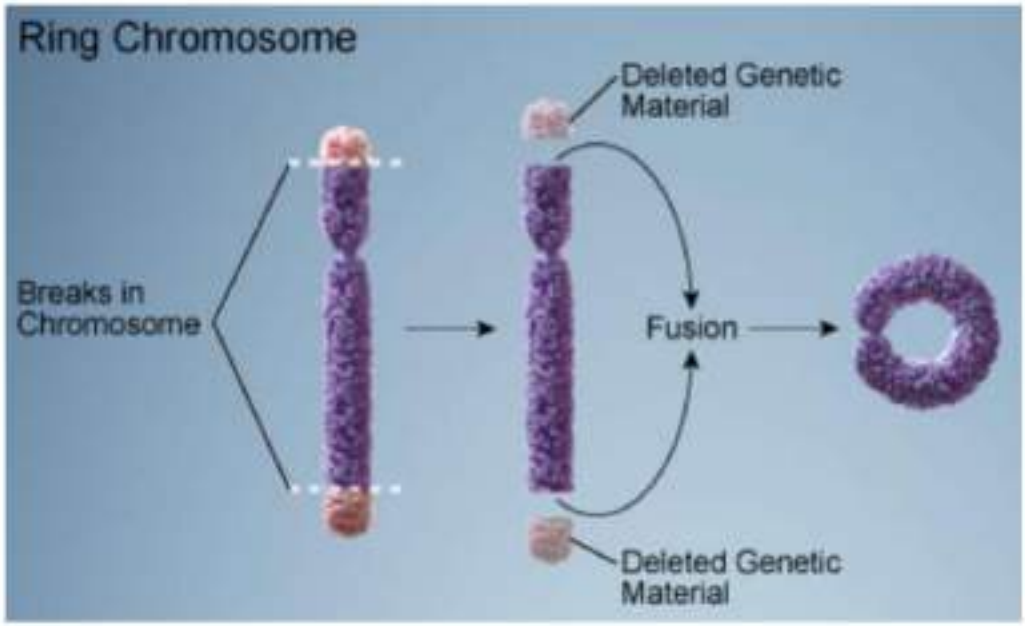


Figure 8: Structure of ring chromosome.

HUMAN CHROMOSOMES

Of the 46 chromosomes in a normal human somatic cell, 44 are autosomes and 2 are sex chromosomes. The autosomes are designated as pairs 1–22. The numbers are assigned in descending order of the length, size, and centromere position of each chromosome pair. In a normal female the sex chromosomes are XX, and in a normal male, they are XY.

Until the advent of certain specialized staining techniques, arbitrary identification of individual chromosome pairs was based on the size and position of the centromere (4). Variability in the centromere position of different chromosomes allowed them to be classified into three basic categories. A chromosome with its centromere in the middle is metacentric, one with the centromere closer to one end is sub-metacentric, and one with the centromere almost at one end is acrocentric.

Based on decreasing relative size and centromere position, a karyotype comprised of seven groups labeled A through G was devised. The X chromosome belonged to the third or "C" group, whereas the Y was often placed separately. Although still used occasionally, these letter group names are now considered obsolete.

Chromosome Banding and Identification

Unequivocal identification of individual chromosomes and chromosome regions became possible with the technical developments of the late 1960s. When chromosome preparations are treated with dilute solutions of proteolytic enzymes (trypsin, pepsin, etc.) or salt solutions (2X SSC) and treated with a chromatin stain such as Giemsa, alternating dark and light stained demarcations called bands appear along the length of each chromosome. The banding patterns produced are specific for each chromosome pair, thus enabling the identification not only of individual chromosomes but also of regions within each chromosome. Methods commonly used to produce these discriminative banding patterns include Giemsa or G-banding,

quinacrine mustard or Q-banding, reverse or R-banding and constitutive heterochromatin or C-banding, each with its own uniqueness. In the United States and Canada, the most frequently used methods for routine cytogenetic analysis are G- and Q-bands, whereas in other countries (France, for example), R-banding is more common. Additional banding methods are occasionally employed to exemplify specific abnormalities or chromosome regions.

THE KARYOTYPE

A karyotype is the characteristic chromosome complement of a eukaryote species. Karyotype descriptions follow certain basic rules. When designating a karyotype, the first item specified is the total number of chromosomes, including the sex chromosomes present in that cell, followed by a comma and the sex chromosomes in that order. Thus, a normal female karyotype is written as 46, XX and a normal male karyotype as 46, XY. The characters are contiguous, without spaces between items. Chromosome abnormalities, when present, follow the sex chromosome designation using abbreviations or symbols denoting each abnormality. These are listed in a specific order: Sex chromosome abnormalities are described first, followed by autosomal changes in numerical order. For each chromosome described, numerical changes are listed before structural abnormalities. The chromosomes are depicted (by rearranging a microphotograph) in a standard format known as a karyogram or ideogram: in pairs, ordered by size and position of centromere for chromosomes of the same size. Karyotypes can be used for many purposes, such as, to study chromosomal aberrations, cellular function, taxonomic relationships, and to gather information about past evolutionary events.

Staining for the study of karyotypes:

The study of karyotypes is made possible by staining. Usually, a suitable dye is applied after cells have been arrested during cell division by a solution of colchicine. For humans, white blood cells are used most frequently because they are easily induced to divide and grow in tissue culture. Sometimes observations may be made on non-dividing (interphase) cells. The sex of an unborn fetus can be determined by observation of interphase cells (see amniotic centesis and Barr body).

Six different characteristics of karyotypes are usually observed and compared:

1- Differences in absolute sizes of chromosomes. Chromosomes can vary in absolute size by as much as twenty-fold between genera of the same family: Lotus tenuis and Vicia faba (legumes), both have six pairs of chromosomes (n=6) yet V. faba chromosomes are many times larger. This feature probably reflects different amounts of DNA duplication.

2- Differences in the position of centromeres. This is brought about by translocations.

3- Differences in relative size of chromosomes can only be caused by segmental interchange of unequal lengths.

4- Differences in basic number of chromosomes may occur due to successive unequal translocations which finally remove all the essential genetic material from a chromosome, permitting its loss without penalty to the organism (the dislocation hypothesis). Humans have one pair fewer chromosomes than the great apes, but the genes have been mostly translocated (added) to other chromosomes.

5- Differences in number and position of satellites, which (when they occur) are small bodies attached to a chromosome by a thin thread.

6- Differences in degree and distribution of heterochromatic regions.

Heterochromatin stains darker than euchromatin, indicating tighter packing, and mainly consists of genetically inactive repetitive DNA sequences.

A full account of a karyotype may therefore include the number, type, shape and banding of the chromosomes, as well as other cytogenetic information. Variation is often found:

- between the sexes
- between the germline and soma (between gametes and the rest of the body)
- between members of a population (chromosome polymorphism)
- geographical variation between races
- mosaics or otherwise abnormal individuals

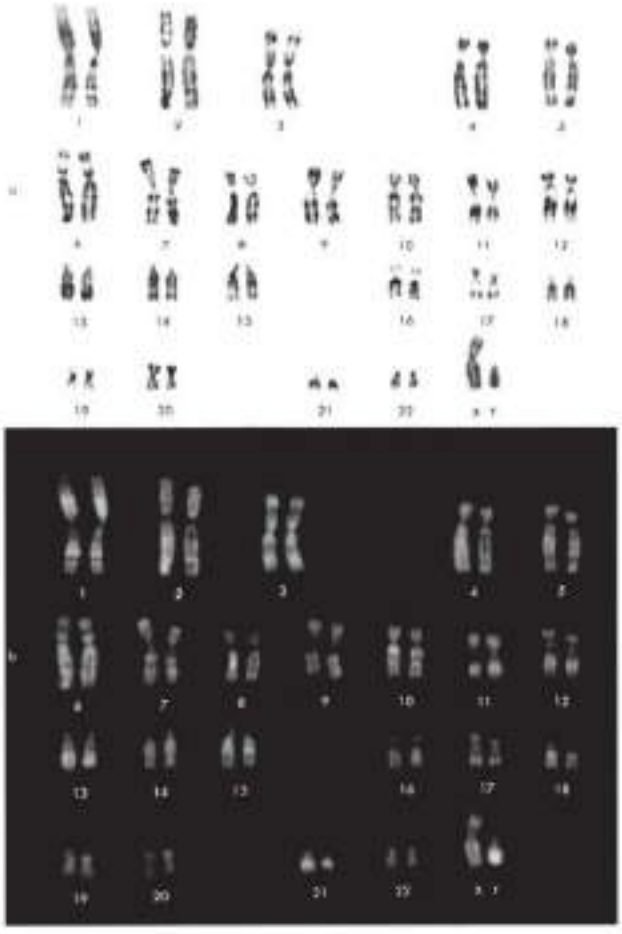


Figure 9: Normal 46, XY male karyotype. Characteristic G-band pattern **(a)** and fluorescent Q-banding.

THE IDEOGRAMS

Ideograms are a schematic representation of chromosomes. They show the relative size of the chromosomes and their banding patterns. A banding pattern appears when a tightly coiled chromosome is stained with specific chemical solutions and then viewed under a microscope. Some parts of the chromosome are stained (G-bands) while others refuse to adopt the dye (R-bands). The resulting alternating stained parts form a characteristic banding pattern which can be used to identify a chromosome. The bands can also be used to describe the location of genes or interspersed elements on a chromosome.

Below is an ideogram of each human chromosome. Next to the known schematic representation a chromosome was added that was rendered from DNA data. The G-bands, areas with proportional more A-T base pairs, are normally colored black in schematic representations. To compare the schematic ideograms the rendered chromosomes, were colored the A-T bases black and the G-C bases white. Blue areas in the rendered chromosomes identify bases not known yet.

The results are interesting, when comparing the schematic ideograms with the rendered chromosomes from our project, a significant conformance can be seen. Most black areas on the left side show also black areas on the right side and white areas are also white on the "digital" chromosomes.

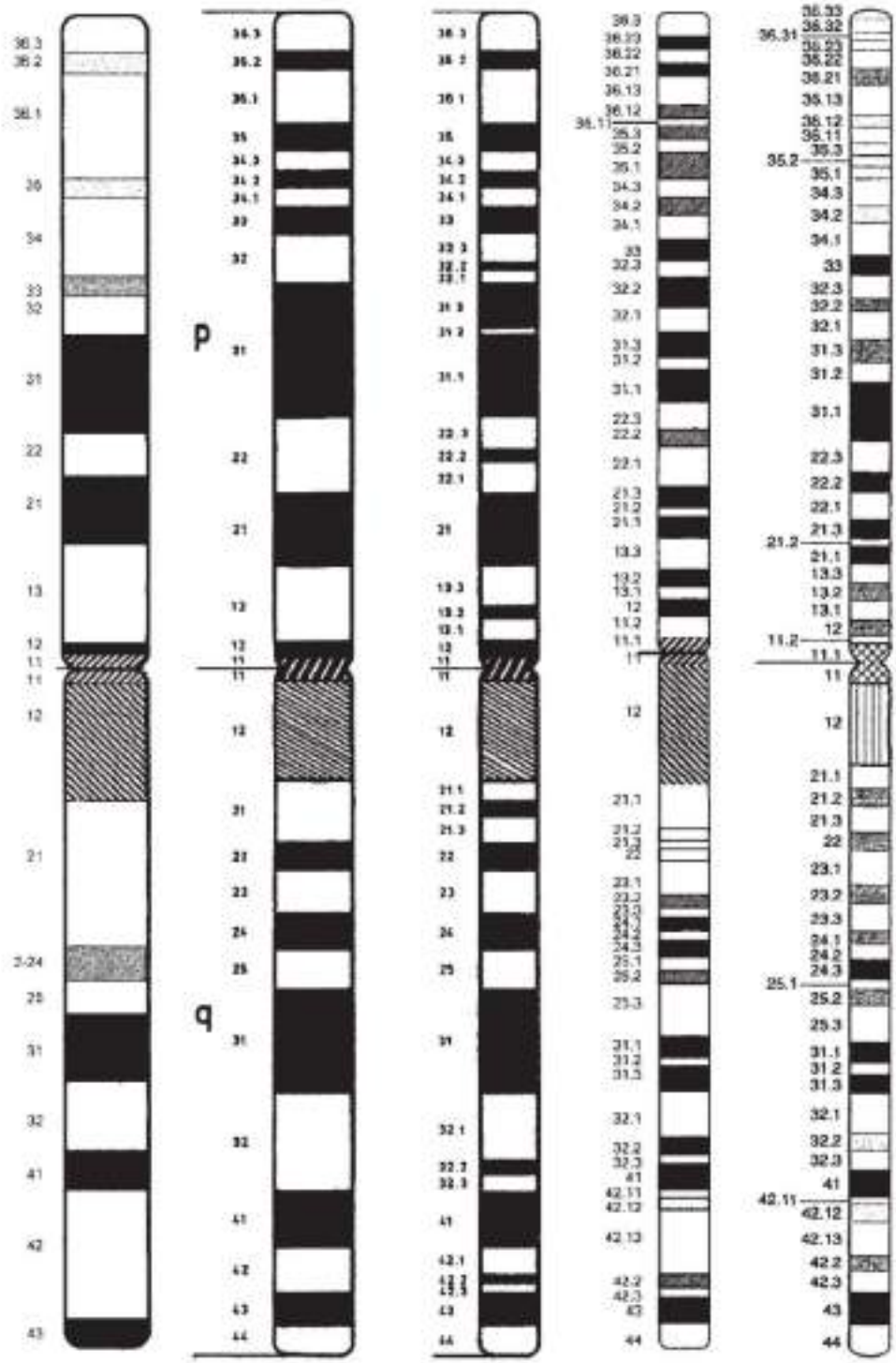


Figure 10: Ideogram showing the G-banding pattern for normal human chromosomes at different band resolutions. From left to right these are 300, 400, 550, 700, and 850 bands.

CHROMOSOMAL ABERRATION

Two major kinds of chromosomal aberrations:

I- Structural aberrations

Chromosomes may undergo changes. This is called chromosomal mutation or chromosomal aberration. The change may occur either in structure of the chromosomes or in the number of chromosomes. Based on these, the chromosomal aberrations are grouped into two major kinds- variation in structure and variation in number.

Variation in chromosome structure

These are four kinds of variations in the structure of chromosomes.

1. Deletion
2. Duplication
3. Inversion
4. Translocation

(1) Deletion

Definition: It is an "intra-chromosomal aberration" in which, an interstitial or terminal chromosomal segment is lost. That is, some genes are deleted. Based on which it is called inter calary or terminal deficiency.

Cytological effect: In deletion heterozygotes, "deletion loop" occurs during pairing of homologous chromosomes. The portion of the normal chromosome homologous to the deficient segment bulges out.

Genetic effect: When a segment of a chromosome is absent, some genes are also absent. If these lost genes are physiologically important, deletion leads to death of the organism. Deficiencies produce unique phenotypic effects in *Drosophila*. The characters such as bended, delta, gull, minute, and notch are associated with some deletions in chromosomes.

In human beings, **Cri-du-chat' syndrome** is characterized by a mewing cat like cry during infancy, widely spaced eyes physical and mental retardations. This 'Cri- du-chat' syndrome is caused by a deletion in the short arm of 5th chromosome.

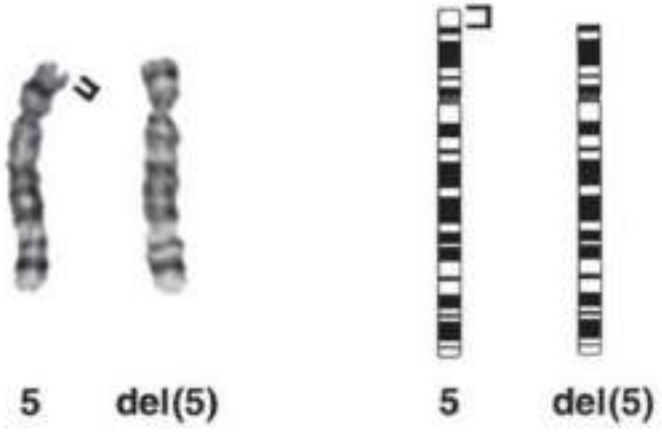


Figure 11: A terminal deletion involving the distal short arm of chromosome 5 [del(5)(p15.3)]. Patients with similar deletions are said to have cri du chat or cat cry syndrome because of the characteristic cat like cry present in many during infancy.

Figure 12: An interstitial deletion involving the long arm of chromosome 13 [del(13)(q21.3q33)].

(2) Duplication

Definition: It is an 'intrachromosomal aberration' in which a segment isrepresented two or more times in a chromosome.

Cytological effect: During meiotic pairing of heterozygotes, the chromosome with duplicated segment forms a loop to maximize the juxtaposition of similar segments of homologous chromosomes.

Genetic effect: The duplications are not lethal. The unusual dosage of genes can be investigated. Duplications are useful in evolution of new characters without loss of original traits. Relocations of chromosomal materials, due to duplication,results in an altered phenotype. This is called position effect.

Position effect

The position effect is an altered phenotype due to relocation of chromosomal material. A fly homozygous for Bar eye has four 16 A segments, two in each chromosome. A fly heterozygous for ultra-Bar has also four 16 A segments – one in the normal chromosome and three in the duplicated number of 16A segments, they areexpected to be similar in phenotype. But the flies homozygous for ultra-Bar (BB/+) produced smaller size eyes.

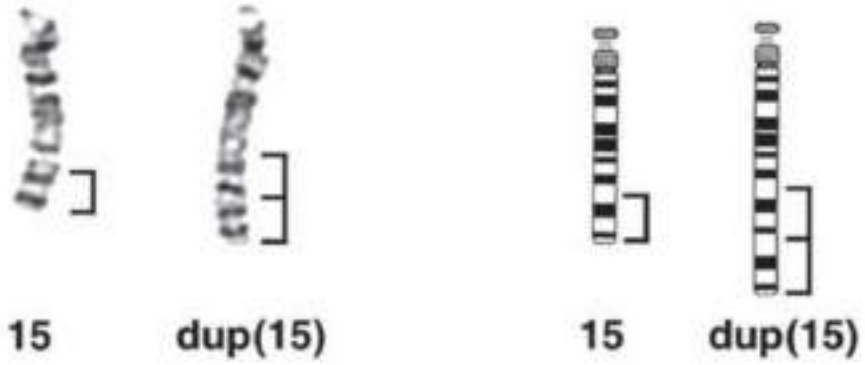


Figure 13: A duplication involving the distal long arm of chromosome 15 [dup (15)(q24q26.3)]. This duplication was initially observed in the bone marrow of a patient with mental retardation and leukemia. By obtaining a peripheral blood karyotype, we were able to demonstrate that the duplication was constitutional and apparently unrelated to his leukemia.

(3) Inversion

Definition: It is an intrachromosomal aberration. Inversions occur when a part of chromosome become detached, turn through 180° and reinserted in such a way that the genes are in reverse order.

Inversions are of two kinds:

(i) Pericentric inversion: The inverted segment includes centromere.

(ii) Paracentric inversion: The inverted segment does not include centromere. Centromere lies outside the inverted portion.

Origin of inversion: A chromosome may form a loop. Breakages occur at the pointof intersection. When the sticky ends unite with new partners, inversion results.

Cytological effect: In inversion heterozygote the part of the uninverted chromosome corresponding to the inversion forms a loop. A similar loop is formed by the inverted section of the homologous chromosome but in reverse direction.

Genetic effects: Paracentric inversion produce dicentric and acentric chromosomes. Pericentric inversion produce duplications and deficiencies. Inversion acts as cross over suppressor and inversion maintains heterozygosity from generation to generation.

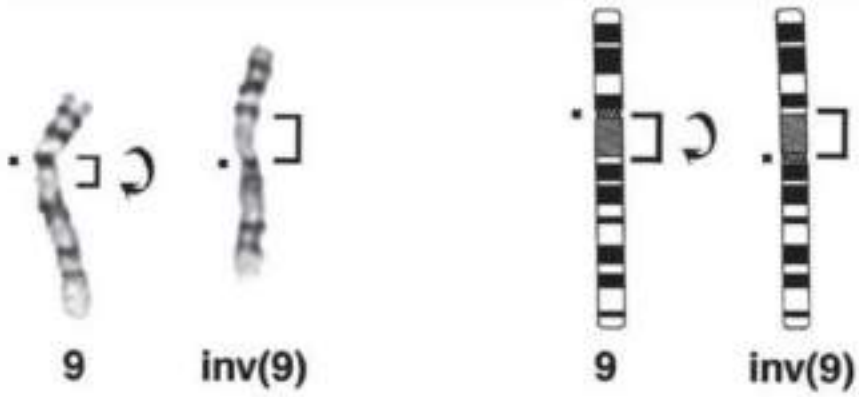


Figure 14: This benign inversion of chromosome 9 [inv(9)(p11q13)] represents a pericentric inversion with breakpoints in both chromosome arms.

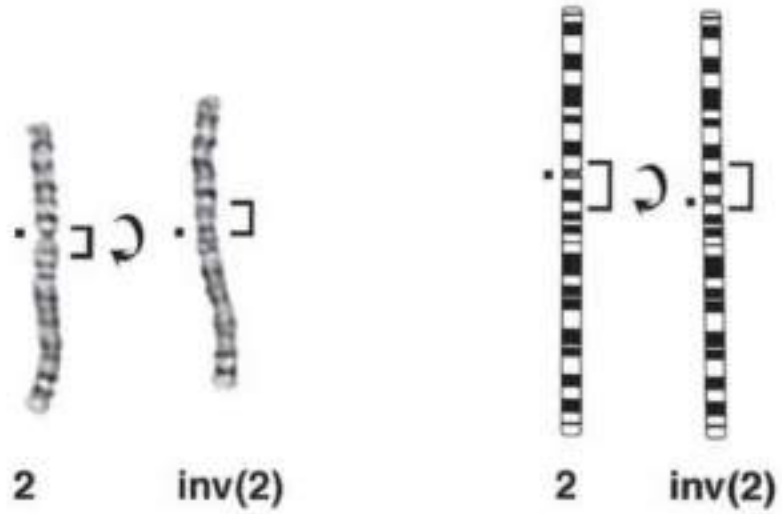


Figure 15: Although this recurring pericentric inversion [inv(2) (p11q13)] is considered to be benign, individuals who carry this inversion might have a slightly increased risk for miscarriages.

(4) Translocation

Definition: It is an inter-chromosomal aberration where in exchange of chromosomal a segment occurs between non-homologous chromosomes.

Cytological effect: In the translocation heterzygote, pairing of homologous chromosomal segments is affected by a cross-shaped configuration. This cross opens out into a ring as chiasma terminalizes. The meiotic products are of three kinds (i) normal, (ii) balanced and (iii) unbalanced.

Genetic effect: Translocation gives three kinds of genetic effects.

i. Translocations alter the linkage relationships of genes.

ii. Heterozygotic translocation causes semi sterility because most of the gametes fail to receive full, balanced complement of genes required for viable development.

iii. The phenotypic expression of a gene may be modified when it is translocatedto a new position.

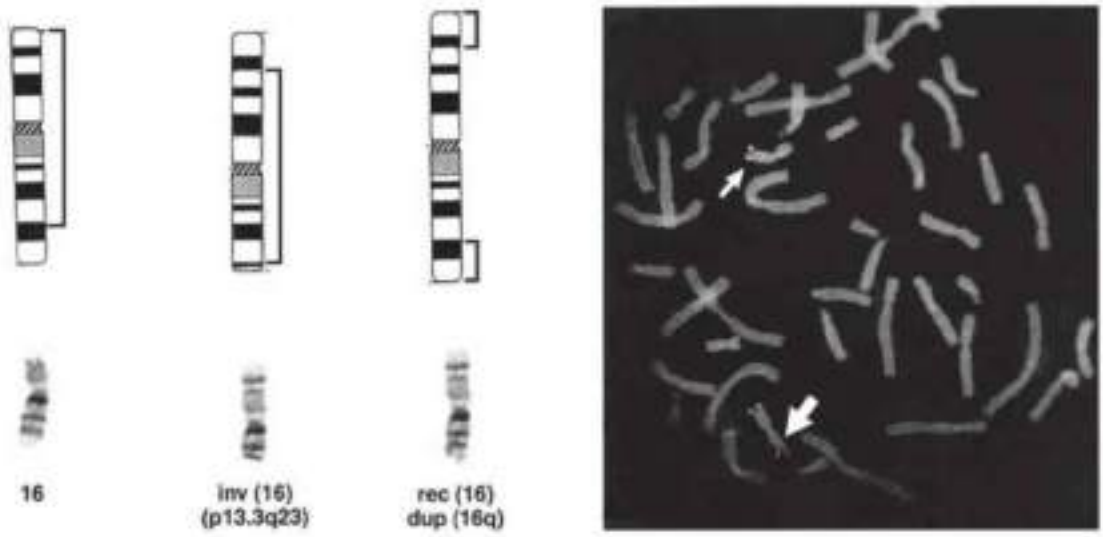


Figure 16: A normal 16, an inverted chromosome 16, and a recombinant chromosome 16 [rec(16) dup(16q) inv(16)(p13.3q23)] resulting from recombination within the inversion loop of the parental inversion carrier. The recombinant chromosome 16 is missing the material distal to the short arm breakpoint and contains a duplication of the material distal to the breakpoint within the long arm.

II- Numerical aberrations(Variation in Chromosome number):

Variation in number of chromosomes is called **ploidy**. A set of chromosomes present in an organism is called **genome.** In a genome, each type of chromosome is represented only once. Most of the sexually reproducing plant species are diploids i.e.,have two set of chromosomes. Any change in the chromosome number from the diploid condition is referred to as heteroploidy. The heteroploidy is of two types namely, aneuploidy and euploidy. The variation in number may involve any chromosome or in entire sets.

▪ Aneuploidy

Loss or gain of one or more particular chromosomes occur within a set is called aneuploidy. The aneuploidy organism bears irregular number of chromosomes. Aneuploidy arises due to non-disjunction. Aneuploidies are of three types.

Types of Aneuploids:

Types	Genomic constitution
Monosomic	2n-1
Double monosomic	2n-1-1
Nullisomic	2n-2
Trisomic	2n+1
Double trisomic	2n+1+1
Tetrasomic	2n+2
Pentasomic	2n+3

1. Monosomic

A monosomic is an individual that lacks one chromosome of the normalcomplement of somatic cells (2n-1). If the lost chromosome is one that is not absolutely essential for the organism, it may survive but if the lost chromosome isvery important, the organism may not survive.

2. Nullisomic

A nullisomic is an individual that lacks both members of one specific pair of chromosomes (2n-2). A nullisomic diploid does not survive. However, a nullisomic polyploidy (hexaploidy wheat 6x-2) may survive but exhibit reduced vigour and fertility. Nullisomic analysis helps to identify genes with specific chromosomes in a polyploidy species.

3. Polysomic

An individual having either single or one pair of extra chromosome in the diploid complement is known as polysomics. Polysomics are called as hyperploids. Polysomics are of two types (i) trisomics and (ii) tetrasomics.

(i) Trisomics

A trisomic is an individual with one chromosome more than the normal

complement of the somatic cells (2n+1). In general the extra chromosome does not produced so striking effect as a missing one. In wheat, trisomics (2n=43) are nearly indistinguishable from normal plants. Trisomics give rise to two kinds of gamets *i.e.*, one kind with 'n' chromosomes and other with 'n+1' chromosomes. Trisomics are more stable genetically than monosomics.

(ii)Tetrasomics

Addition of two chromosomes of one pair or two different pairs is known as tetras my and such individuals are called as tetrasomics.

Use of aneuploidy

i. Aneploids are extremely useful in several genetic studies.
ii. They are useful to determine the phenotypic effects of loss or gain of different chromosomes.
iii. Aneuploids have been used to produce chromosome substitution lineswhich give information on the effects of different chromosomes of avariety.
iv. They are used to produce alien addition and alien substitution lines which are useful in gene transfers from one species into another.
v. Aneuploid analysis permits the location of a gene as well as of a linkage group of a specific chromosome.

Aneuploids in Human beings

I-Down's syndrome

It is due to trisomic condition of 21st chromosome. It is also called Mongolian idiocy. Affected individuals are mentally deficit and physically retarded, broad face and flat stubby nose.

II- Kline felters syndrome (44+XXY)

It is due to trisomic condition of sex chromosome. The individual is male withXXY

Chromosome. The individuals with this syndrome have defective dev elopement of testis, famine character like Enlarged breast, under- developed body hair, presence of one barr body in the body cells.

III- Turners' syndrome

It is due to monosomic condition of sex chromosome. The individual is female with 44 autosomes and one'X' chromosome. The female individual is without menstrual cycle. No barr body is present in body cells.

The origin of aneuploids

i. Spontaneous
ii. Meiotic irregularities
iii. Triploid individuals
iv. Translocation heterozygote

Use of aneuploids in crop improvement

a. Aneuploids are useful tools for locating the genes on a specific chromosome. Monosomics and nullisomics are used for this prupose.
b. Monosomics are also used in interspecific gene transfer *ie* monosomics are used in transferring chromosomes with a desirable genes from one species to another.
c. Aneuploids are used for developing alien addition and alien substitution lines in various crops.
d. Primary trisomics are useful in identification of chromosomes involved in translocations.

- **Polyploidy**

These are variation that involves entire set of chromosomes. In Euploids the chromosome number is an exact multiple of the basic or genomic number. Euploids are differ in multiple of n or x.

Types	Genomic formula
Monoploids	n
Diploid	2n
Triploid	3n
Tetraploid	4n
Pentaploid	5n
Hexaploid	6n

1. Monoploid

The monoploid organisms have one set of chromosomes or one genome (n) in the nuclei of their body cells. The monoploids are often weak and sterile. Monploids differ from haloids which carry half or gametic chromosome number (n). In true diploid species, both monoploid and haploid chromosome number is the same (n =x) thus a monoploid can be a haploid but all haploids cannot be monoploids.

2. Diploids

Normal diploids are known as disomic. They have regular bivalent pairingduring meiosis. Diploids also have disomic genetics with two alleles at each locus.

3. Polyploids

Polyploids refer to any organism in which the number of chromosomes sets exceeds two *i.e.*, an organism with more than two set of chromosomes or genome. They have larger cells than diploids. These larger cell sizes contribute to largerplant size and

higher yield. Polyploids have generally larger, thicker and darker green leaves bigger flower, fruits than the diploids. In each genus, there is an optimum level of polyploidy beyond which growth may be depressed with increasing number of chromosomes. (eg) triploid (3n).

There are two types of Polyploids.

I.Autopolyploid

In autoployploids the multiple sets of chromosomes are identical (eg). Genome is identical with each other. Autopolyploids arise by abnormal mitosis and meiosis and induced artificiallyby colchicines.

Auto triploid - 3x

Auto tetraploid- 4x

Auto hexaploid- 6x

(eg). Banana 2n=3 x=33, Groundnut 2n=4x=40; Sweet potato 2n=6x=90 andPotato 2n=4x=48

II.Autotriploid

The triploid organisms have three sets of chromosomes. A triploid may originate by the union of a monoploid gamete (n) with a diploid gamete (2n). Since anautotriploid remains sterile and cannot produce seeds, it has great commercial value inproducing seedless varieties of economic plantgs. Eg. Seedless water melon.

III.Autotetraploid:

The organisms with four genomes (4n) in the nuclei of their somatic cells are called tetraploids. They arise due to somatic doubling of chromosome number. Doubling is accomplished by either spontaneously or it can be induced by chemicals such as colchicines.

Morphological and Cytological features of polyploids are

i. Larger in size than diploids
ii. Generally more vigorous than diploids
iii. Slower in growth and late in flowering
iv. Polyploids may have reduced fertility than diploids

Role of polyploids and their evoluation

- About 1/3 of angiosperms are polyploids. These suggest that polyploids have significant role in the evolution of crop species.
- Allopolyploids have contributed great extent in the evolution of plants than auto polyploids.
- The identification of diploid parental species is primarily based on paising between the chromosome of diploid and the allotetraploid species.
- Allopolyploids combine the genome of different species, hence the resulting individuals differ from progenitor.
- Evolution is a slow process; but due to allopolyploids new species originatevery quickly.
- Polyploids have wider adaptation to different environmental condition than diploids.

GENETIC DISORDERS

Disease	Gene/Defect	Inheritance	Clinical Features
Achondroplasia	Fibroblast growth factor receptor 3 (FGR3) – constitutively active (gain of function)	Autosomal dominant (normal parents can have an affected child due to new mutation, and risk of recurrence in subsequent children is low)	Short limbs relative to trunk, prominent forehead, low nasal root, redundant skin folds on arms and legs
Cystic Fibrosis	Cystic fibrosis trans-membrane regulator (CFTR) – impaired chloride ion channel function	Autosomal Recessive (most common genetic disorder among Caucasians in North America)	Pancreatic insufficiency due to fibroticlesions, obstruction of lungs due to thick mucus, lung infections (Staph, aureus, Pseud. aeruginosa)
Duchenne Muscular Dystrophy	Dystrophin (DMD) -deletions	X-linked recessive	Gradual degeneration of skeletal muscle, impaired heart and respiratorymusculature
Hypercholesterolemia	LDL receptor (commonly)	Autosomaldominant (haploinsufficiency)	Impaired uptake of LDL, elevated levels of LDL cholesterol, cardiovascular disease and stroke. Symptoms more severe in homozygous individuals
Fragile X Syndrome	(FMR1) – CGG trinucleotide repeat expansion in 5' untranslated region of the gene (expansion occurs exclusively in the mother)	X-linked dominant (females less severely affected) Inheritance characterized by anticipation	Disorder shows anticipation (female transmitters in succeeding generations produce increasing numbers of affected males) Boys have long faces, prominent jaws, large ears, and mentally retarded.
Gaucher's Disease	B-Glucosidase	Autosomal recessive	Lysosomal storage disease characterized by splenomegaly, hepatomegaly, & bone marrow infiltration. Neurological symptoms are not common
Glucose 6-phosphate dehydrogenase deficiency	Glucose 6-phosphate dehydrogenase	X-linked recessive (Prominent among individuals of Mediterranean and African descent)	Anemia (due to increased hemolysis) induced by oxidizing drugs, sulfonamide antibiotics, sulfones (e.g.dapsone), & certain foods (e.g. fava beans)

Hemochromatosis	Unknown gene on the short arm of chromosome 6	Autosomal recessive (Incidence ~0.3% in Caucasoid population. Women less affected due to increased iron loss through menstruation)	Enhanced absorption of dietary iron with accumulation of abnormal, pigmented, iron-protein aggregates (hemosiderin) in visceral organs. Cirrhosis, cardiomyopathy, diabetes, skin pigmentation, and arthritis.
Holoproencephaly	Sonic Hedgehog (SHH)	Autosomal dominant (haploinsufficiency?)	Malformation of the brain (no or reduced evidence of an interhemispheric fissure), dysmorphic facial features, mental retardation
Huntington Disease (Huntington Chorea)	Huntingtin (HD) – CAG repeat expansion within exon 1 (expansion occurs in father)	Autosomal dominant (gain-of- function mutation) Shows anticipation	Disorder is characterized by progressive motor, cognitive & psychiatric abnormalities. Chorea – nonrepetitive involuntary jerks is in 90% of patients
Klinefelter Syndrome	47,XXY males	50% of cases due to errors in paternal meiosis I	Sterile males with long limbs, small genitalia, breast development, and feminine body contours, and learning disabilities
Marfan Syndrome	Fibrillin-1 gene (FBN1) encodes a microfibril forming connective tissue protein	Autosomal dominant (dominant negative effect)	Abnormalities of the skeleton (disproportionate tall stature, scoliosis), heart (mitral valve prolapse, aortic dilatation, dissection of the ascending aorta), pulmonary system, skin (excessive elasticity), and joints (hypermobility). A frequent cause of death is congestive heart failure.
Myoclonic Epilepsy with Ragged Red Fibers (MERRF)	Mitochondrial DNA mutation in the tRNAlys gene	Maternal transmission, heteroplasty	Age of onset varies depending on fraction of mutant mitochondrial DNA inherited. Symptoms include myopathy. (disease takes its name from abnormal histological appearance of skeletal muscle biopsies), dementia, myoclonic seizures, ataxia, and deafness

Myotonic Dystrophy	A protein kinase gene (DMPK) – CTG repeat expansion in 3' untranslated region of the gene	Autosomal dominant Shows anticipation	Disorder shows anticipation. Muscle weakness, cardiac arrhythmias, cataracts and testicular atrophy in males. Children born with congenital form have a characteristic open triangle-shaped mouth
Neurofibromatosis I	Microdeletion at 17q11.2 involving the NF1 gene	Autosomal dominant	The disorder is characterized by numerous benign tumors (neurofibromas) of the peripheral nervous system, but a minority of patients also show increased incidence of malignancy(neurofibrosarcoma, astrocytoma, Schwann cell cancers and childhood CML – chronic myelogenous leukemia)
Osteogenesis Imperfecta	Either of the genes encoding the α1 or α2 chains of type I collagen	Usually autosomal dominant (null mutations result in haploinsufficiency, missense mutations often produce a dominant negative effect	Null mutations produce a milder form of the disease. Missense mutations that act in a dominant negative manner are often perinatal lethal. The disorders are associated with deformed, under mineralized bones that are subject to frequent fracture.
Phenylketonuria	Usually due to a mutation in Phenylananine hydroxylase (PAH)	Autosomal recessive	Mental retardation, if untreated, possibly due to inhibition of myelination and disruption of neurotransmitter synthesis. Detectable by newborn screening and treatable
Polycystic Kidney Disease	Mutations in either polycystin-1 (PKD1) or polycystin-2 (PKD2) gene	Autosomal dominant (disease appears to follow a "two- hit model", requiring the loss of both alleles of PDK1 or PDK2 for the disease to be evident.	Heterozygous individuals are predisposed to polycystic kidney disease because they are likely to loose the second good copy of the gene during their lifetime. Multiple renal cysts, blood in urine, end-stage renal disease and kidney failure.

Prader Willi/Angelman (PWS/AS)	Deletion of the PWS region and AS gene located at 15q11- q13. Can also be caused by uniparental disomy involving chromosome 15	Complex Parent of origin effects due to genomic imprinting.	Inheriting the deletion through the mother gives rise to Angelman syndrome, which is characterized by short stature, severe mental retardation, spasticity, seizures, and a characteristic stance. Inheriting the deletion from the father produces the more common Pader-Willi syndrome, which is characterized by obesity, excessive and indiscriminate gorging, small hands, feet, hypogonadism and mental retardation. In rare cases, uniparental disomy involving chromosome 15 produces PWS when both copies are inherited from the mother and AS when both copies are inherited from the father.
Sex Reversal	Variety of causes	Various	See Thompson & Thompson, Medical Genetics, 6th ed.
Tay-Sachs Disease	B-Hexosaminidase (A isoenzyme (HEXA)	Autosomal recessive (common among Jew of Eastern European ancestry and French Canadians).	Hypotonia, spasticity, seizures, blindness, death by age 2. An early indication is a cherry red spot on the retina. (Incidence greatly reduced by screening)
Thalasemias		Autosomal Recessive	Severe anemia
Turner Syndrome	45,X females	Usually due to a paternal error in sex chromosome transmission	Although usually lethal in utero, the defect poses little risk to survival in infants that do come to term. Short stature, webbed necks, broad chest with widely spaced nipples, and sterility. Infants show evidence of lymphedema in fetal life. Intelligence is normal.

DNA AND RNA AS A GENETIC MATERIAL

Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) the principal genetic materials of living organisms are chemically called nucleic acids. Nucleic acid especially the DNA, a universal genetic material of most of the organisms, is having all the features required to be a good genetic material. DNA is a macromolecule and is a helically twisted double chain of poly deoxyribonucleotides.

In prokaryotes it occurs in nucleoid and as plasmids, both are double stranded circular DNA. In Eukaryotes most of the DNA is found in chromatin of nucleus. It is linear. Some small quantitative of DNA are found in mitochondria and plastids which is generally double stranded and circular RNA also acts as genetic material in majority of plant viruses.

Features of DNA to act as genetic material:

- Genetic material can store information used to control both the development andmetabolic activities of cell.
- It should be chemically stable so that it can be replicated accurately during cell division.
- It should be transmitted for generations.
- It should be able to undergo mutations providing genetic variability required for theevolution.

(A) Structure of DNA

Nucleic acid (DNA or RNA) first called nuclein by a Swiss chemist Friedreich Miescher (1869) as he removed nuclei from pus cells and isolated DNA i.e., "nuclein" from it. Nucleic acid (DNA or RNA) are macromolecules composed of repeating subunit called nucleotides.

Constitution of a nucleotide:

- A phosphate groups.
- A five-carbon sugar (ribose in RNA and deoxyribose in DNA).
- A cyclic nitrogen containing compound called a base (purines and pyrimidines).

Most commonly DNA occurs as a double helix. The two spiral strands of DNA are collectively called DNA duplex. Two separate and anti-parallel chains of DNA are wound around each otherin a right-handed helical manner. The DNA double helix comes to have two types of alternate grooves major and minor with the sugar phosphate backbone on the outer sides. The bases paired by hydrogen bonding are stocked on each other.

Chemical Composition of DNA

Deoxyribonucleotides (monomer) of DNA are composed by three different types of chemicals.

(1) **Phosphoric acid (H_3PO_4)** has three reactive (-OH) groups of which two are involved informing sugar phosphate back bone of DNA.

(2) **Pentose sugar ($C_5H_{10}O_4$)** - DNA contains 2'-deoxy-D-ribose, hence the name deoxyribose.

(3) **Nitrogen bases**- DNA contained four different nitrogen bases (**A, G, C & T**). These four bases are grouped in to two classes on the basis their chemical structure.

(a) Purine bases - DNA has two types of purines (adenine and guanine). Each purine is a type of nitrogen base having a double ring structure (i.e., 9 member double rings with nitrogen at 1, 3, 7 and 9 positions).

Some of the common names of these bases reflect the circumstances of their discovery. Guanine, for example, was first isolated from guano (bird manure),

and thymine was firstisolated from thymus tissue.

(b) Pyrimidine bases: DNA has two types of pyrimidine bases (cytosine and thymine). Each pyrimidine is a type of nitrogen containing base having a single ring structure (i.e.6 member rings with nitrogen at 1 and 3 positions).

Nucleosides: A nitrogenous base with a molecule of deoxyribose sugar (without phosphate group) is known as nucleosides. In nucleic acids, the nitrogen bases are covalently attached to the 1'-position of a pentose sugar ring with the help of glycosidic bond.

Nitrogen base + sugar = nucleoside.

- Adenine + deoxyribose = deoxyadenosine
- Guanine + deoxyribose = deoxyguanosine
- Cytosine + deoxyribose = deoxycytidine
- Thymine + deoxyribose = deoxythymidine

Nucleotides- A nucleotide is formed of one molecule of deoxyribose sugar, one molecule of phosphoric acid and anyone of the nitrogen base. Phosphoric molecule is attached to the 5^{th} – carbon atom of deoxyribose ring with the help of phosphoester bond.

Nucleosides + phosphoric acid = nucleotides

Different nucleotides of DNA are as follows:

(1) Adenine + deoxyribose + phosphoric acid = deoxyadenylic acid or deoxyadenylate / dAMP

(2) Guanine + deoxyribose + phosphoric acid = deoxyguanylic acid or deoxyguanylate / dGMP

(3)Cytosine + deoxyribose + phosphoric acid = deoxycytidylic acid or deoxycytidylate /dCMP

(4)Thymine+deoxyribose+phosphoric acid = deoxythymidylic acid or deoxythymidylate / dTMP

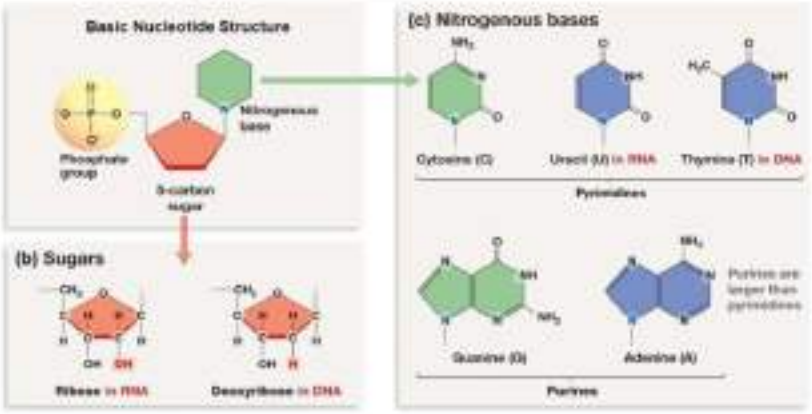


Figure 17: Chemical structure of nucleic acids structure.

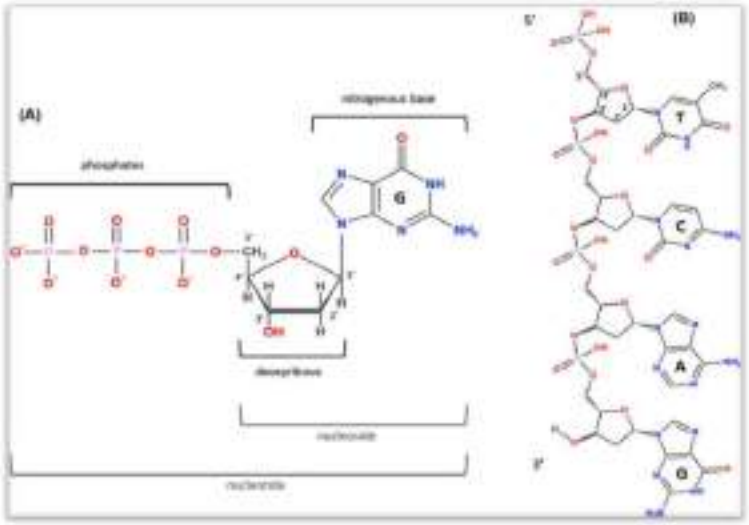


Figure 18: (a) Progressive formation of nucleoside to nucleotide (from lower to higher energy compounds, (b) Backbone of DNA. (The backbones are formed by 3 -to-5 phosphodiester linkages).

Nitrogen base	Nucleoside (Nitrogen base + sugar)	Nucleotide (nucleocide +phosphate gp.)
Adenine (A)	A+S= Adenosine	Adenylic acid adenosine monophosphate (AMP)
Guanine (G)	G+S= Guanosine	Guanylic acid Guanosine monophosphate (GMP)
Thyamine (T)	T+S = Thyamidine	Thyamidylic acid Thyadine monophosphate (TMP)
Cytosine (C)	C+S = Cytidine	Cytidylic acid Cytidine monophosphate (CMP)

Figure 19: Nitrogen bases, their respective nucleosides, and nucleotides of DNA.

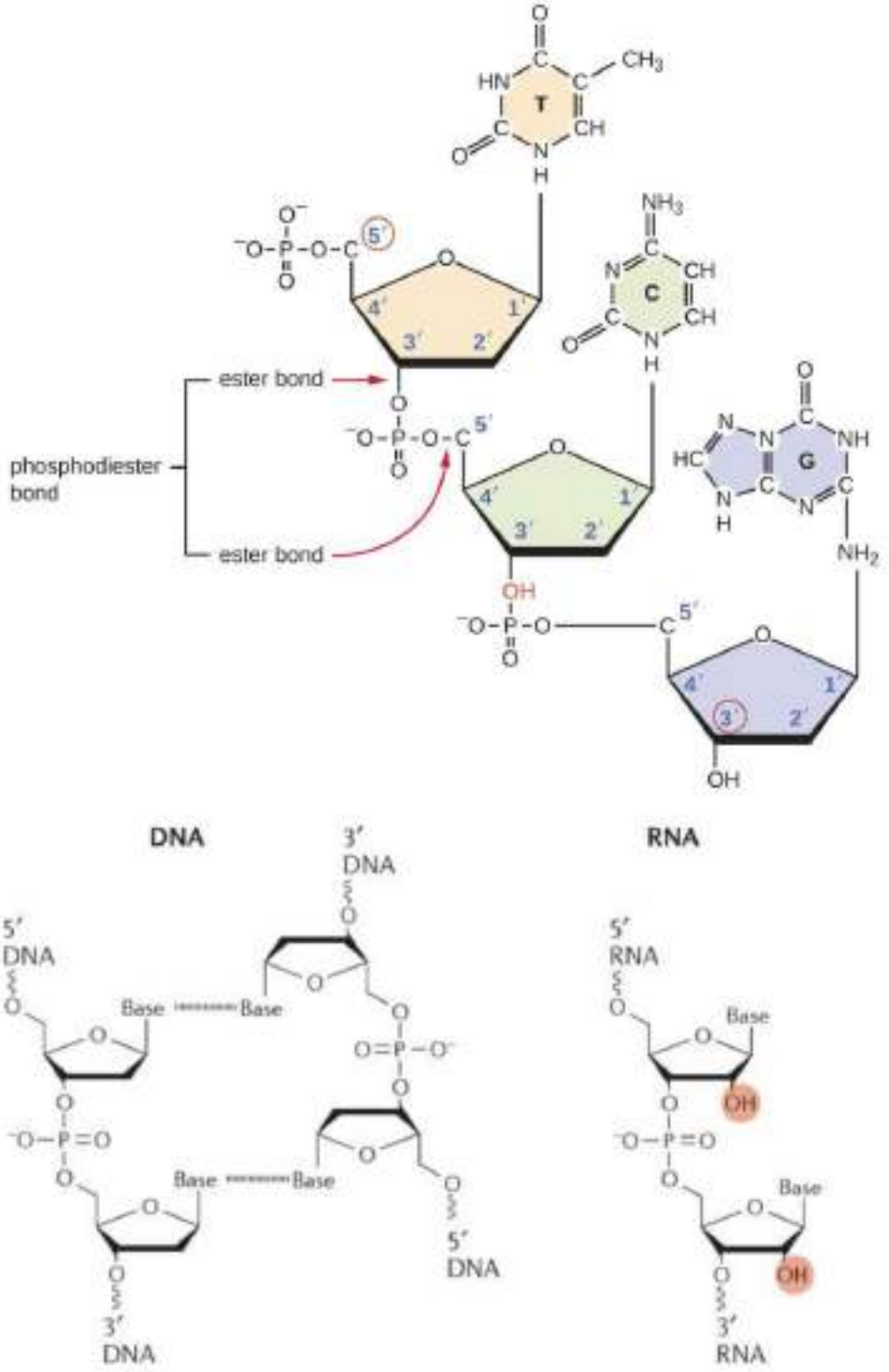


Figure 20: Chemical constituents of a nucleotide and DNA & RNA chain formation.

Watson and Crick Double Helix Model of DNA:

The structure of DNA was deduced by American J. D. Watson and F.H.C. Crick in 1953 for which they received the Nobel Prize in 1962. Their double- helix model of DNA structure model is widely accepted. Their double helix model of DNA was based on the data and informationgiven by so many workers like E. Chargaff, M.H.F. Wilkins, R. Franklin and their coworkers. Main contributions in deducing this model were of: Chargaff's rule, Franklin's X-ray diffraction patterns and Kornberg's results.

Chargaff's rule- In 1940's Erwin Chargaff analyzed base content of DNA using new chemical techniques and their observations and generalizations were called as Chargaff 's rule. Chargaff's rule strongly suggested that thymine and adenine as well as cytosine and guanine were present in DNA, always bonded to each other by H-bonds and shows some fixed inter relationship.

- The proportion of A always equals that of T, and the proportion of G always equals thatof C or A = T and G = C.
- The amount of A, T, G, and C in DNA vary from species to species but A+T/G+C =constant for a particular species.

Franklin's X-ray diffraction patterns- Watson and Crick made use of the data of x-ray crystallographic of DNA structure from the studies of M.H.F. Wilkins, R. Franklin, and their coworkers. According to their data, DNA was a highly ordered, multiple stranded structure with repeating sub structure spaced every 3.4A° along the axis of the molecule.

Korenberg's results: Korenberg and his associates tried to synthesize DNA in a medium free of DNA but in the presence of enzyme **DNA polymerase** and nucleotides-the building blocks of DNA. They found that in a DNA free medium

with all necessary compounds DNA synthesis does not occur but the same happens i.e., DNA synthesis starts only when some DNA was added as a primer to the same medium.

The important features of their model of DNA are:

a. Two helical polynucleotide chains are coiled around common axis, where the backbone isconstituted by sugar phosphate and the bases project inside.
b. The polynucleotide chains run in opposite directions. It means, if one chain has the polarity 5'P→ 3'OH, the other has 3'OH → 5'P.
c. The two chains are held together by hydrogen bonds between their bases. Three hydrogen bonds occur between cytosine and guanine (C≡G) and two hydrogen bonds between adenine and thymine (A=T).
d. The diameter of the helix is $20A^0$ and bases are separated by 3.4 A^0 along the helix axisand related by a rotation of 36^0.
e. The helical structure repeated after 10 residues on each chain, and intervals of 34 A^0

Functions of DNA:

1- DNA is genetic material which able to store information used to control both thedevelopment and metabolic activities of cells.
2- DNA can be replicated accurately during cell division and transmitted for generations.
3- Crossing over during meiosis produces natural recombination of DNA which is passed on to next generation to produce variants in all sexually reproducing organisms.
4- DNA able to undergo mutations providing genetic variability required for evolution

5- Differentiation of various body parts is due to differential functioning of specific parts of DNA.

6- Developmental stages occur in the life cycle of an organism by an internal clock of DNA functioning.

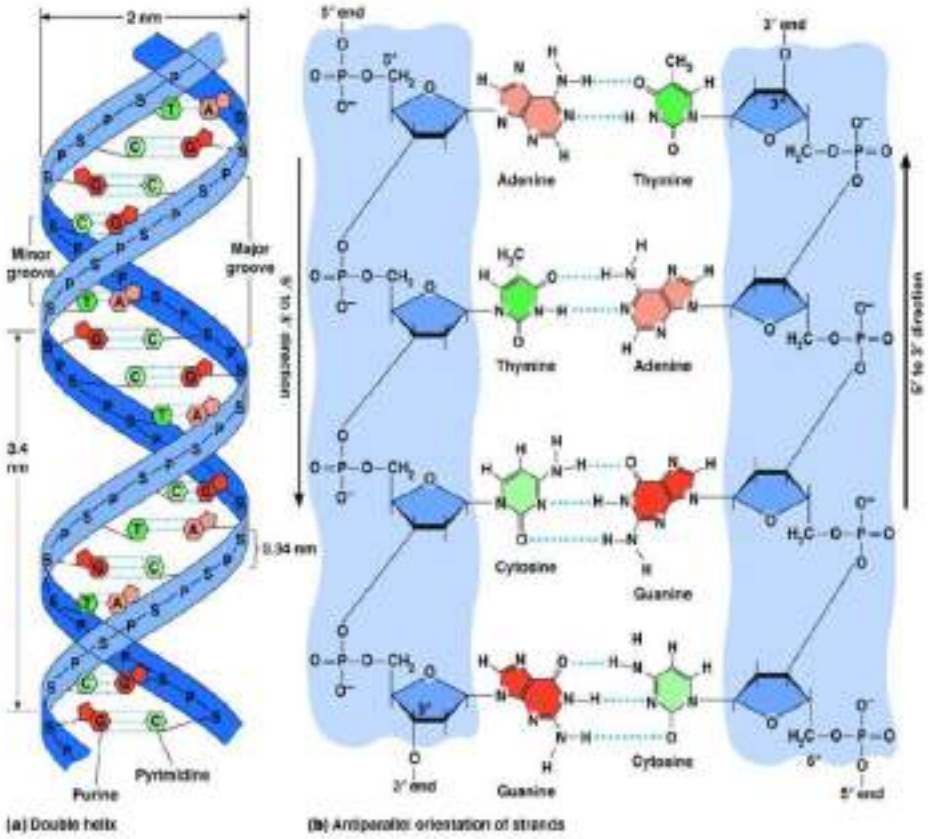


Figure 21: Watson and Crick Double Helix Model of DNA.

Replication of DNA:

Replication is the process of formation of carbon copies on DNA. DNA functions as its own template. DNA replication is an autocatalytic function of DNA. During DNA replication the weak hydrogen bonds between nitrogen bases of the nucleotides separate so that the two polynucleotide chains of DNA separate and uncoil. The chains thus separated are complementary to one another. Each stand acts as a template and makes its own complimentary copy over it so that the new formed DNA duplex has one parental stand and one newly formed strand. This method of formation of new daughter DNA molecules is called semi-conservative method of replication.

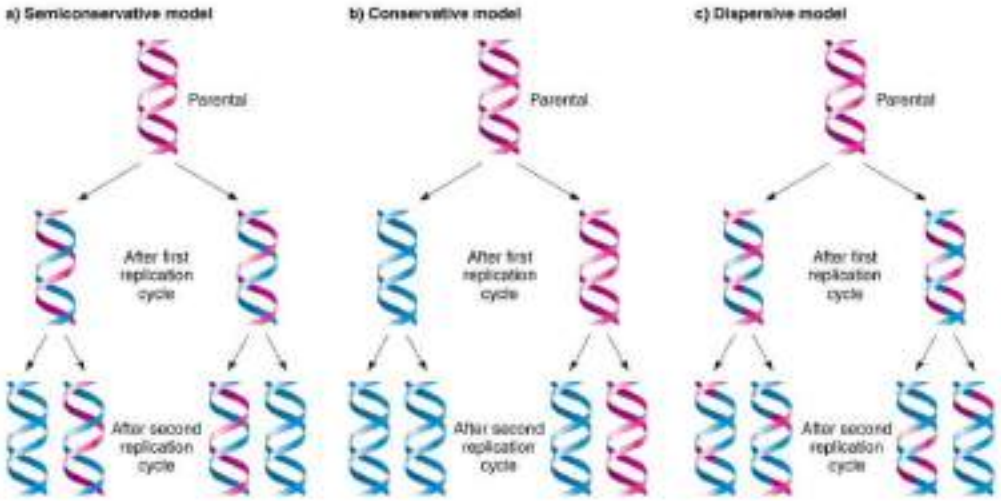


Figure 22: DNA replication.

Mechanism of DNA Replication:

DNA replication is the process of copying a DNA molecule and involves following four major steps:

1. Initiation of DNA replication.
2. Unwinding of helix.
3. Formation of primer strand.
4. Elongation of new strand.

1. **Initiation of DNA replication-** Replication is regulated by the rate of initiation. Replication of DNA in E. coli always begins at a definite site called **origin of replication**. The ***E. coli,*** origin of replication lies within the genetic locus **'ori'** and is bond to the cell membrane. 'Ori' contains four 9bp binding sites for the initiator protein (DnaA-ATP). The helicase DnaB (or mobilepromoter) binds and extends the single-stranded region for copying.

2. **Unwinding of helix-** Unwinding of DNA molecule into two strands results in the formationof Y shaped structure called **replication fork.** Due to unwinding positive super coiling has to be relieved by the **enzyme topoisomerase or DNA Gyrase**.

3. **Formation of Primer strand:** As the newly formed replication fork displaces the parental lagging strand, a mobile complex called a **primosome**, which includes the DnaB, Helicase and DNA primase help in the synthesizes **of RNA primers**. Both leading and lagging strand primers are elongated by **DNA polymerase III.** Need of primer is there to facilitate the action of DNA polymerase III as this enzyme cannot initiate the process but can add activated deoxyribonucleotides to the 3' OH end of primer.

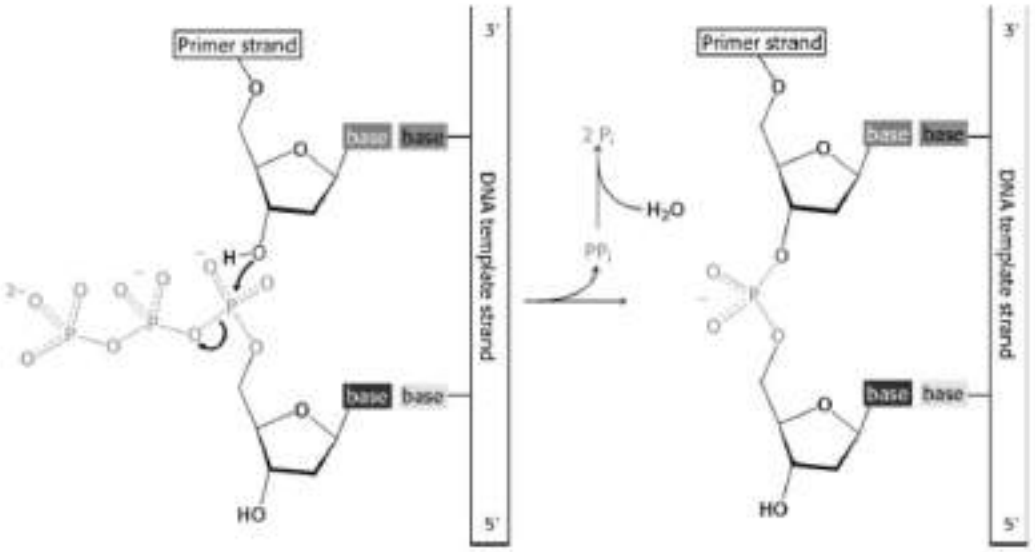


Figure 23: DNA Replication (Phosphodiester Bridge is catalyzed by DNA polymerases).

4. **Elongation of new strand:** after the formation of primer strand, DNA replication occurs in 5' → 3'direction and complementary deoxyribonucleotides are added only to the free 3'OH end of the primer. A dimer of DNA polymerase III elongates both leading (3' → 5') and lagging strands. The leading strand shows continuous replication while the lagging strand shows discontinuous replication. These short pieces of DNA replicated against lagging strand are known as **Okazaki fragments.** Okazaki fragments are 1000-2000 nucleotides long in prokaryotes. A separate RNA primer is used for the synthesis of each Okazaki fragments which, after replacing the RNA primers from deoxyribonucleotides, are later joined together with the help of **DNA ligase** or **DNA synthetase** forming a continuous lagging strand. Hence DNA replication is semi- discontinuous as the leadingstrand is synthesized continuously and lagging strand is formed discontinuously in short pieces join later.

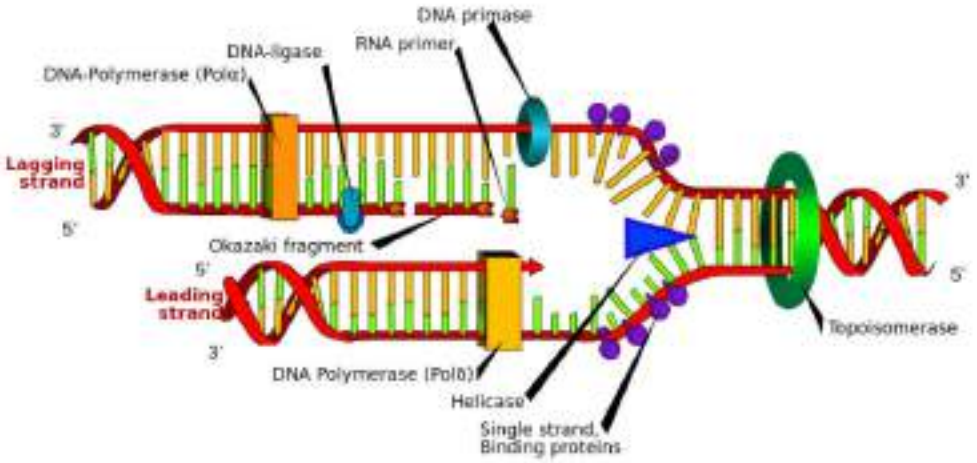


Figure 24: DNA Replication. During DNA replication, a number of different enzymes work together to pull apart the two strands so each strand can be used as a template to synthesize new complementary strands. The two new daughter DNA molecules each contain one pre-existing strand and one newly synthesized strand.

Recombinant DNA:

The tools and technologies of molecular biology for breaking and rejoining DNA sequences from two or more different organisms are known as DNA recombinant technologies. These modified DNA fragments are called recombinant DNA. A recombinant DNA molecule is a vector in which the desired DNA fragment has been inserted to enable its cloning in an appropriate host. This is achieved by using specific enzymes (restriction enzymes) for cutting the DNA into suitable fragments and then for joining together the appropriate fragments by ligation.

(B) Structure of RNA:

RNA is generally involved in protein synthesis but in majority of plant and some animal virusesit also acts as genetic material. There are two major types of RNA:

1. **Genetic RNA-** H. Fraenkel-Conrat showed that RNA present in **Tobacco Mosaic Virus** isits genetic material and this RNA is responsible for the infection in tobacco plant.

2. **Non- genetic RNA-** Prokaryotes and Eukaryotes where genetic information is contained in the DNA molecule, functions of such cells are performed by a different kind of nucleic acids called non- genetic ribonucleic acid. Non-genetic RNA is synthesized on DNA template. Suchnon genetic RNAs can be of many types like mRNA, r RNA, & t RNA.

Chemical structure of RNA:

RNA is single stranded polyribonucleotide. Each ribonucleotide is made of:

- Phosphoric acid- H_3PO_4
- Ribose sugar- $C_5H_{10}O_5$
- Nitrogen base- Adenine (A), Guanine (G), Cytocine (C) and Uracil (U)

Many ribonucleotides join with each other by phosphor-ester bonds to make a linear chain of polyribonucleotide's. The chain will remain straight under all conditions in mRNA, may fold randomly in r-RNA or specifically to form t-RNA.

Types of RNA:

The RNA is of following three major types: t RNA, mRNA, and r RNA.

(1) Transfer RNA or t-RNA:

It is also called **soluble or s-RNA**. There are over 100 types of t-RNA. t-RNA is the smallest RNA with 70-85 nucleotides and sedimentation co-efficient of 4S. It is about 10-15% of the total weight of tRNA of the cell. Each tRNA has a corresponding **anticodon** that can recognize the codon on mRNA and exhibit high affinity for specific activated amino acids combine with them and carry them to the site of protein synthesis.

Robert Holley (1965) and his colleagues reported the complete nucleotide sequence of alanine tRNA of yeast. R. Holley (1965) first of all proposed a **clover leaf model for yeast tRNA^{ala}. Cloverleaf structure-**. Five parts or arms of cloverleaf structure:

(1) Acceptor stem or arm - this is a region of the tRNA which acts as a site of attachment for the appropriate amino acid. It is also called **amino acid carrier arm**. It is formed by seven regular Watson & Crick base pairs between the 5' and 3' end of the tRNA. The **3' terminal end** of all tRNA is **always CCA-OH**. It is not base- paired and is the site of attachment of the amino acid. The amino acid is covalently bound through an ester linkage between the carboxyl group of the amino acid and the 3' hydroxyl group of the ribose of the tRNA.

(2) Anti-codon loop or arm - The anti-codon loop contains the three-nucleotide sequence that is complementary to the codon of mRNA to which it corresponds. It consists of a total of 7 unpaired bases, three of which constitute the anti-codon. With this site tRNA attaches to mRNA and helps in the transport of amino acids to the site of protein synthesis

(3) DHU loop or D loop or arm - The DHU loop is composed of three or four base pairs. It is depending on the species of tRNA. It is also variable in size containing 8 to 12 unpaired bases. The D-loop helps in binding of amino-acyl synthetase. It has modified bases called dihydrouridine hence named so.

(4) T φC loop or arm- is named so because of the presence of triplet sequence of pseudouridine (φ). It acts as ribosome recognize arm, help in determining the site of ribosome (A, P or E site) where the tRNA must come and attach during translation.

(5) The extra arm- is variable in nucleotides composition and is lacking entirely in some tRNA

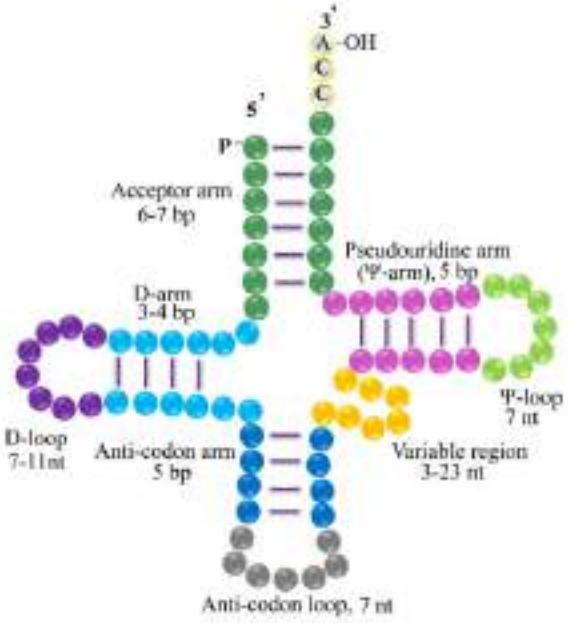


Figure 25: t-RNA structure.

Functions of t-RNA:

The tRNA plays important role in protein synthesis. T-RNA picks up a specific amino acid from the cytoplasm carries it to the site of protein synthesis and attaches itself to ribosome in accord with the sequence specified by mRNA. It transmits its amino acid to the polypeptide chain. In protein synthesis tRNA acts an adaptor molecule which is meant for transferring amino acids to ribosomes for synthesis of polypeptides. There are different tRNAs for different amino acids. Codons are recognized by anticodons of tRNA. They hold peptidyl chains over the mRNAs.

(2) Messenger RNA or (mRNA):

The structure of mRNA:

m-RNA is **always single stranded** having normal bases like A, G, U and C along with only afew unusual, substituted bases. There is never base pairing in mRNA. It functions as a templatefor protein synthesis it carries genetic information from DNA to a ribosome and helps to assemble amino acids in their correct order. Each amino acid in a protein is specified by a set of three nucleotides in the mRNA called **codons.** Both prokaryotic and eukaryotic mRNA contains three primary regions:

a) **5' untranslated region (5'UTR) -** the 5' untranslated region is a sequence of nucleotides atthe 5' end of the mRNA that does not code for the amino acid sequence of a protein. In **prokaryotic** (bacterial cell) mRNA contains a consensus sequence called the **Shine-Dalgarno sequence** (5'AGGAGGU3'), which serves as **the ribosome binding site during translation**, itis formed of approximately 7 nucleotides upstream of the first or start codon. Eukaryotic mRNA has no such equivalent sequences in its 5' untranslated region. This is the sequence of the mRNA extending from the 5' end of the mRNA to the initiation codon. It is not translated into polypeptide sequence. It has a **function**

analogous to the function of a promoter on a gene. Itwill direct the binding of the ribosome to the initiation codon.

b) **Protein coding region**- this region comprises the codon that specify the amino acid sequence of the protein. This region begins with a start codon and ends with a stop codon. This region has 3 regions namely initiation codon, coding region, stop codon.

- **Initiation codon-** it is always **AUG** and codes for a **methionine.** This is the triplet codon at which polypeptide synthesis begins. All polypeptides are synthesized with an amino terminal methionine.
- **Coding region-**this is the sequence of mRNA that contains **the consecutive triplet codons** that direct polypeptide synthesis. This region starts from the start codon and continue up to the stop codon. The coding region is often referred to as the open reading frame or ORF.
- **Stop codon-**this is the triplet codon that signals the **termination of translation**. There are three possible stop codon sequences **UAA, UAG, UGA**. Stop codons have no corresponding tRNA or amino acid.

c) **3' Untranslated region (3'UTR)-**This region of mRNA is the 3' un-translated region, a sequence of nucleotides at the 3'end of mRNA that is not translated into protein. This is the nucleotide sequence downstream from the stop codon. It extends from the stop codon to the 3' end of the mRNA. It does not code for amino acid sequence. It may function in stabilizing the mRNA. In eukaryotes it is transcribes as hnRNA which is converted into functional mRNA in the cytoplasm by removing introns (intervening sequences) and joining together exons (expressible sequences).

For the convenience the mRNA structure can be summarized as:

1. **Cap**- at 5' end, has methylated structure, does not translate.
2. **Noncoding region-1**- has 10-100 nucleotides, rich in U and A bases, does not translate.
3. **The initiation codon**- AUG, codes for methionine amino acid
4. **The coding region**-about 1500 nucleotides on an average, translate proteins.
5. **Termination codon**- either of UAA, UAG or UGA i.e., present, helps in termination of translation.
6. **Noncoding region-2**- made of 50-150 nucleotides, does not translate, has sequence likeAAUAAA.
7. **Poly(A) sequence**- 200-250 A nucleotides, does not translate, makes tail of mRNA.

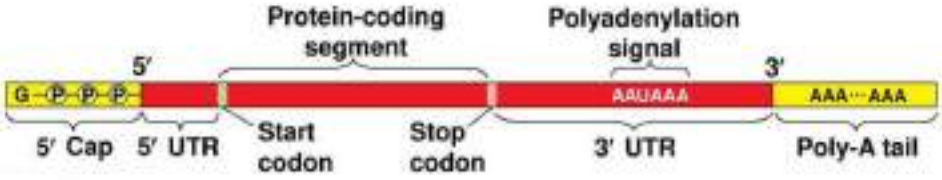


Figure 26: mRNA showing different regions.

Functions of m-RNA:

m-RNA carries coded information to be translation into polypeptide. It directly takes part in protein synthesis in a cell. In some viruses having RNA as genetic material, it may undergo reverse transcription to from compact genes which are used in genetic engineering. The phenomenon also occurs in nature and has added certain genes in the genomes.

(3) Ribosomal rRNA (r-RNA):

Ribosomal, stable, or insoluble RNA constitutes the largest part (up to 80%) of the total cellular RNA. It was reported by **Kuntz.** It is found primarily in the cytoplasm as well as organelle. In prokaryotes it is transcribed from ribosomal DNA which is

a part of nuclear DNA but in eukaryotes ribosome is formed on nucleolar DNA. The genetic instruction contained in mRNA is translated into the amino acid sequences of polypeptides only with the help of ribosomes. Thus, ribosomes play an integral part in the transfer of genetic information from genotype to phenotype. R-RNA is most stable type of RNA.

Structure and processing of ribosome RNA:

It forms about 80% of the total cellular RNA. r- RNA consists of a single stranded RNA which gets twisted over itself in certain regions due to complementary base pairing. R-RNA strand unfold on heating and refold on coiling. It is one the most stable RNA among all types of RNAs. R-RNA and ribo-proteins constitute ribosomes.

In **eukaryotes** 4 types of rRNAs found are **28s, 18s, 5.85s, and 5s.** In the nucleolus of eukaryotes, RNA polymerase-I transcribes the rRNA genes, which usually exit in tandem repeats to yield a long, single pre-rRNA which contains one copy each of the 18s, 5.8s and 28s sequences. Various spacer sequences are removed from the long pre-rRNA molecule by a seriesof specific cleavages. Many specific ribose methylations take place directed by small ribonucleoprotein particles (snRNPs) and the mature rRNA molecule fold and complex with ribosomal proteins. RNA pol. III synthesizes the 5srRNA from unlinked genes (Figure 90).

Functions of r-RNA:

r-RNA binds to protein molecules and give rise to ribosomes. 3'end of 18s rRNA (16s in prokaryotes) has unpaired nucleotides complementary to those of region or m-RNA, it is the site where ribosomes bind to mRNA during translation. 5s rRNA and surrounding protein complex provide binding site for tRNA.

Important features of RNA:

- RNA is copied from one strand of the double helix called the template strand.
- RNA differs from DNA in that it is single stranded, has uracil inserted of thymine and has ribose sugar instead of deoxyribose ribose.
- Messenger RNA (mRNA) carries the genetic information that specifies a particular amino acid sequence of protein synthesized.
- mRNA bases constitute codons, each codon is made of three consecutive bases in a row.
- rRNA joins certain proteins to form ribosomes. Ribosomes physically support the other structures involved in protein synthesis, and some rRNA catalyses' formation of peptide bonds.
- tRNA is clover leaf-shaped and connects mRNA codon to an amino.
- In prokaryotes, RNA is translated as soon as it is transcribed while in eukaryotes, RNA is often altered (or modified) before it is actively translated.
- mRNA gains a modified nucleotide cap and a poly A tail.
- Many genes have intervening sequences called introns, which are not transcribed and cutout from the mRNA. The protein encoding sequences in mRNA, exons, are thenreattached. Ribozymes are small RNAs with catalytic activity that can splice introns. They join proteins to form snurps, which associate to form spliceosomes.
- After being processed the RNA must be exported from the nucleus before it is translated.

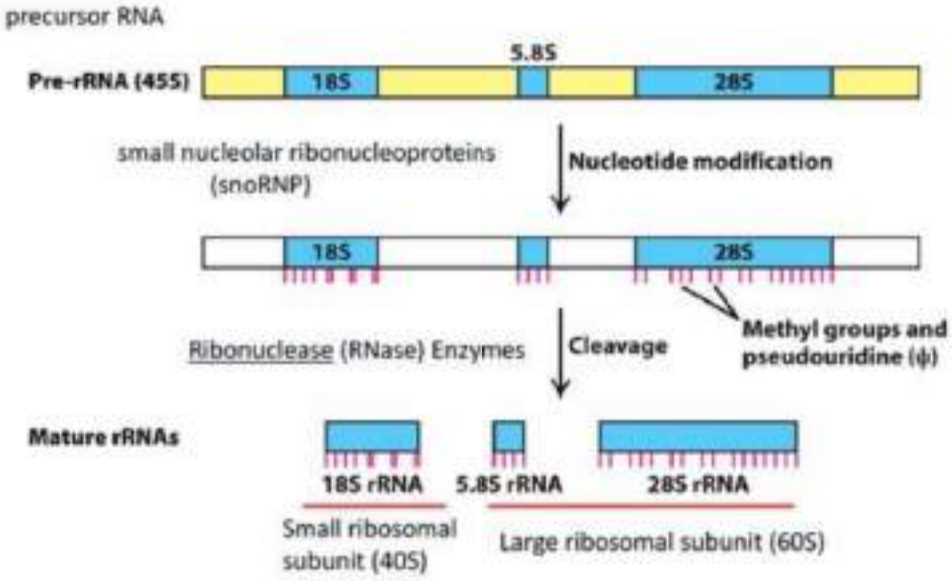


Figure 27: Processing of rRNA in a eukaryotic cell.

GENE EXPRESION AND REGULATION

INTRODUCTION: Each somatic cell in the body generally contains the same DNA. A few exceptions include red blood cells, which contain no DNA in their mature state, and some immune system cells that rearrange their DNA while producing antibodies. In general, however, the genes that determine whether you have green eyes, brown hair, and how fast you metabolize food are the same in the cells in your eyes and your liver, even though these organs function quite differently. If each cell has the same DNA, how is it that cells or organs are different? Why do cells in the eye differ so dramatically from cells in the liver?

Whereas each cell shares the same genome and DNA sequence, each cell does not turn on, or express, the same set of genes. Each cell type needs a different set of proteins to perform its function. Therefore, only a small subset of proteins is expressed in a cell. For the proteins to be expressed, the DNA must be transcribed into RNA and the RNA must be translated into protein. In a given cell type, not all genes encoded in the DNA are transcribed into RNA or translated into protein because specific cells in our body have specific functions. Specialized proteins that make up the eye (iris, lens, and cornea) are only expressed in the eye, whereas the specialized proteins in the heart (pacemaker cells, heart muscle, and valves) are only expressed in the heart. At any given time, only a subset of all of the genes encoded by our DNA is expressed and translated into proteins. The expression of specific genes is a highly regulated process with many levels and stages of control. This complexity ensures the proper expression in the proper cell at the proper time.

Regulation of Gene Expression

For a cell to function properly, necessary proteins must be synthesized at the proper time and place. All cells control or regulate the synthesis of proteins from

information encoded in their DNA. The process of turning on a gene to produce RNA and protein is called **gene expression**. Whether in a simple unicellular organism or a complex multi-cellular organism, each cell controls when and how its genes are expressed. For this to occur, there must be internal chemical mechanisms that control when a gene is expressed to make RNA and protein, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

The regulation of gene expression conserves energy and space. It would require a significant amount of energy for an organism to express every gene at all times, so it is more energy efficient to turn on the genes only when they are required. In addition, only expressing a subset of genes in each cell saves space because DNA must be unwound from its tightly coiled structure to transcribe and translate the DNA. Cells would have to be enormous if every protein were expressed in every cell all the time. The control of gene expression is extremely complex. Malfunctions in this process are detrimental to the cell and can lead to the development of many diseases, including cancer.

Prokaryotic versus Eukaryotic Gene Expression

To understand how gene expression is regulated, we must first understand how a gene codes for a functional protein in a cell. The process occurs in both prokaryotic and eukaryotic cells, just in slightly different manners.

Prokaryotic organisms are single-celled organisms that lack a cell nucleus, and their DNA therefore floats freely in the cell cytoplasm. To synthesize a protein, the processes of transcription and translation occur almost simultaneously. When the resulting protein is no longer needed, transcription stops. As a result, the primary method to control what type of protein and how much of each protein is expressed in a prokaryotic cell is the regulation of DNA transcription. All of the subsequent

steps occur automatically. When more protein is required, more transcription occurs. Therefore, in prokaryotic cells, the control of gene expression is mostly at the transcriptional level.

Eukaryotic cells, in contrast, have intracellular organelles that add to their complexity. In eukaryotic cells, the DNA is contained inside the cell's nucleus and there it is transcribed into RNA. The newly synthesized RNA is then transported out of the nucleus into the cytoplasm, where ribosomes translate the RNA into protein. The processes of transcription and translation are physically separated by the nuclear membrane; transcription occurs only within the nucleus, and translation occurs only outside the nucleus in the cytoplasm. The regulation of gene expression can occur at all stages of the process (Figure 28). Regulation may occur when the DNA is uncoiled and loosened from nucleosomes to bind transcription factors (**epigenetic** level), when the RNA is transcribed (**transcriptional** level), when the RNA is processed and exported to the cytoplasm after it is transcribed (**post-transcriptional** level), when the RNA is translated into protein (**translational** level), or after the protein has been made (**post-translational** level).

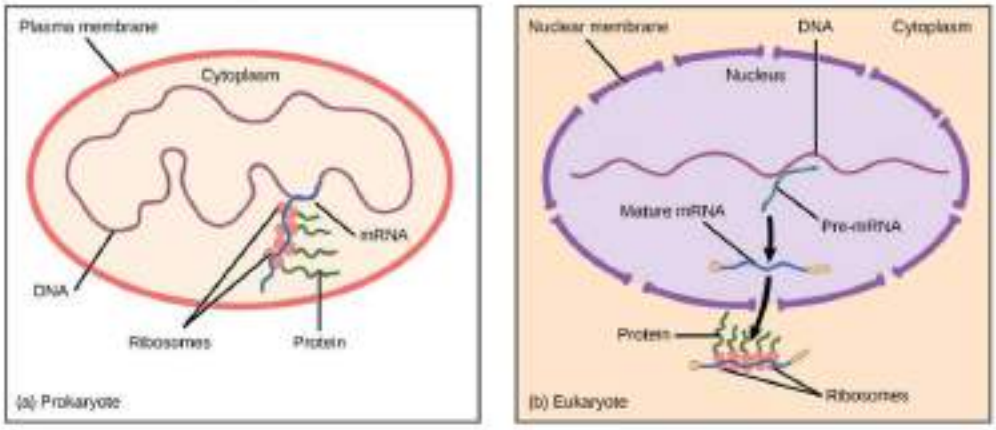


Figure 28: Regulation in prokaryotes and eukaryotes. Prokaryotic transcription and translation occur simultaneously in the cytoplasm, and regulation occurs at the transcriptional level. Eukaryotic gene expression is regulated during transcription and RNA processing, which take place in the nucleus, and during protein translation, which takes place in the cytoplasm. Further regulation may occur through posttranslational modifications of proteins.

PROTEIN SYNTHESIS

The replication of DNA serves to carry genetic information from cell to cell and from generation to generation. This information is translated into protein that determines the phenotype of cell by controlling its biochemical reactions. Protein synthesis is the vital function of the cell where in the genetic information stored in DNA is passed on to RNA, especially mRNA by the process of **transcription.** All the three types of RNA i.e., mRNA, tRNA and rRNA together help in translating the coded information in the form of a polypeptide **(translation).** The linear chain of amino acids translated is the primary protein which undergoes configurationally changes to form secondary, tertiary or quaternary proteins.

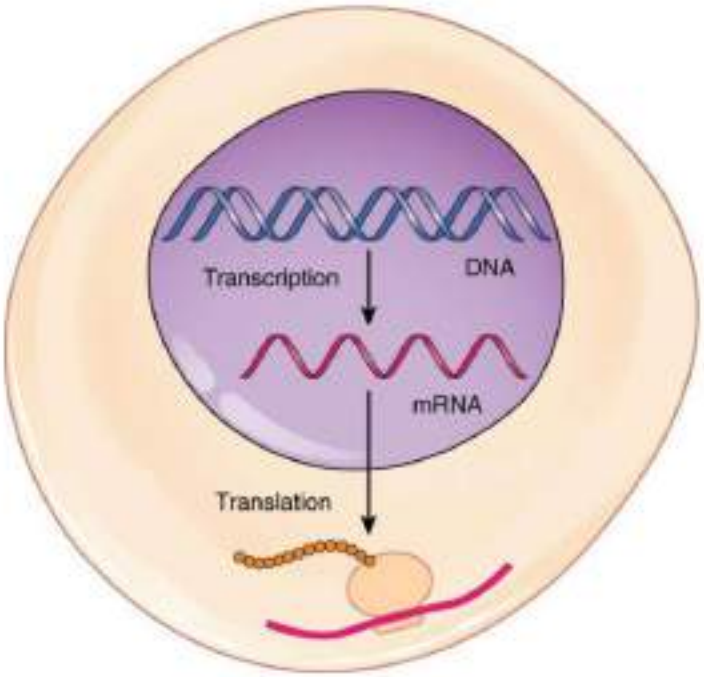


Figure 29: From DNA to Protein (central dogma): Transcription through Translation. Transcription within the cell nucleus produces an mRNA molecule, which is modified and then sent into the cytoplasm for translation.

Protein Synthesis and its Mechanism:

A gene expresses itself by protein synthesis. Protein synthesis is under direct controlof DNA in most cases or else under the control of genetic RNA where DNA is absent. Information for structure of a polypeptide is stored in a polynucleotide chain of DNA orRNA.

In 1958 **F. Crick** proposed that the concept of central dogma, which states that when a particular gene is expressed (control a function or a reactions) its information is copied into another nucleic acid (mRNA) which in turn directs the synthesis of specific proteins. So the central dogma was proposed as unidirectional flow of molecular information from DNA to mRNA and finally to polypeptide. Later a reverse of central dogma was also found in retroviruses. **H. Temin and D. Baltimore** (1970) reported that retro viruses operate a central dogma in reverse manner (inverse flow of information) or teminism inside host cells. This discovery was important in understanding cancer and hence, these two scientists were awarded Nobel Prize.

Genetic RNA of these viruses first synthesizes DNA through reverse transcription. This process is catalyzed by the enzyme reverse transcriptase. DNA then transfers information to messenger RNA which takes part in translation of the coded information to from polypeptide.

Necessary Materials:

(1) **Amino acids-** there are some 20 amino acids and amides which constitute building blocks or monomers of proteins. They are found in the cellular pool or cytoplasm.

The genetic code: it is a set of rules defining how the four-letter code of DNA is translated into the 20-letter code of amino acids, which are the building blocks of proteins. The genetic code is a set of three-letter combinations of nucleotides called codons **(triplet),** each of which corresponds to a specific amino acid or stop signal. The concept of codons was first described by Francis Crick and his colleagues in 1961. There are 64 possible permutations, or combinations, of three-letter nucleotide sequences that can be made from the four nucleotides. Of these 64 codons, 61 represent amino acids, and three are stop signals. Although each codon is specific for only one amino acid (or one stop signal), the genetic code is described as degenerate, or redundant, because a single amino acid may be coded for by more than one codon. It is also important to note that the genetic code does not overlap, meaning that each nucleotide is part of only one codon-a single nucleotide cannot be part of two adjacent codons. Furthermore, the genetic code is nearly universal, with only rare variations reported.

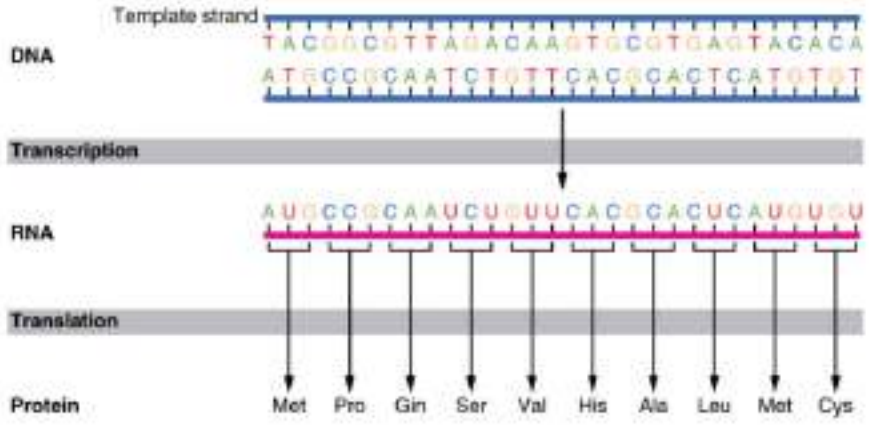


Figure 30: The Genetic Code. DNA holds all the genetic information necessary to build a cell's proteins.

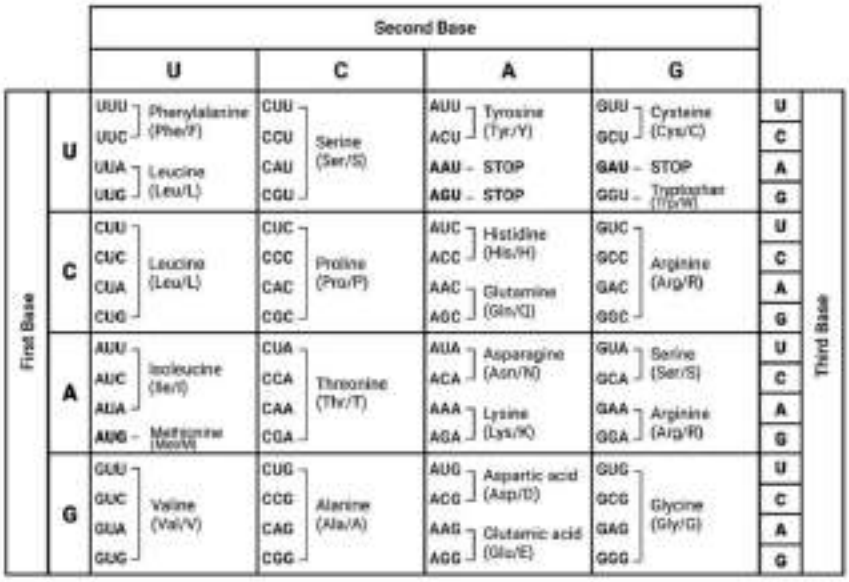


Figure 31: The 20 amino acids formation from 4 nucleotides.

(2) Ribosome- ribosome comprises two subunits which exists as separate subunits prior to the translation of mRNA and contain following sites (Figure 32):

- **P site (peptidyl site or D site- donor site) -** P site is jointly contributed by thetwo ribosomal subunits, most frequently occupied by peptidyl-tRNA or the tRNA carrying growing peptide chain. . The P-site is also referred to as the puromycin sensitive site. Puromycin is an antibiotic which shows similarities with a part of amino acyl-tRNA
- **A site (amino acyl site) -** A site is situated on the larger subunit of ribosome. It faces the tunnel between the two subunits, frequently occupied by amino acyl-tRNA, functions as acceptor for growing protein during peptide bond formation.
- **E-site** – the exit site, the ribosomal site harboring decylated tRNA on transit out from the ribosome.

The different parts of ribosomes, connected with protein synthesis are:

a- **A tunnel**- It lies between the two subunits, acts as a place for mRNA.

b- **The longitudinal groove**-is part of the longer subunits which acts as a passage of newly synthesized polypeptide.

c- **Reactive sites**- P, A and E-site

d- **P-site**- acts as a donor of peptide chain to the newly coming tRNA

e- **A-site**- acts as abinding site for new tRNA with its amino acid for the elongation of **mRNA**- carrying genetic information of DNA into cytoplasm for its translation.

(3) tRNA- to transport the respective amino acids as per their anticodons against thecodons of mRNA.

(4) Enzymes- amino acid activating system **(aminoacyl- tRNA synthetase),** Peptidepolymerase system.

(5) ATP- as energy source.

(6) GTP- for synthesis of peptide bonds.

(7) Soluble protein initiation and transfer factors

(8) Various inorganic cations (K^+, NH_4^+, Mg^{++} or Mn^{++})

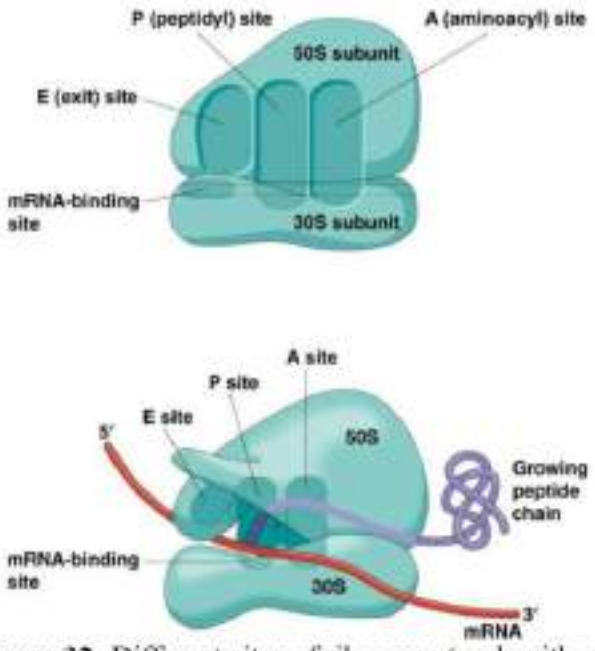


Figure 32: Different sites of ribosome (each with specified function).

Mechanism of protein Synthesis:

Two major steps are involved in protein synthesis are: -

I- Transcription: involving transfer of genetic information from DNA to mRNA.

II- Translation: involving translation of the language of nucleic acid into that of apolypeptide.

I- Transcription process:

The transfer of genetic information from DNA to mRNA in general is known as transcription. The segment of DNA that takes part in transcription is called transcription unit. It has three components:

a) A promoter
b) The structural gene
c) A terminator

a) A promoter- promoter sequences are present upstream (5'end) of the structural genes of a transcription unit. The binding sites for RNA polymerase lies within the promoter sequence. In prokaryotes 10bp upstream from the start point lies a conserved sequence described as 10 nucleotide sequences **TATAAT or "pribnow box"** and 35 nucleotide sequences **TTGACA as "recognition sequence".**

b) The structural gene- structure gene is part of that DNA strand which has 3'-5' polarity as transcription occur in 5'- 3' direction. The strand of DNA that directs the synthesis of mRNA is called **template or non-coding strand**. The complementary strand is called **non-template or coding strand,** it is identical in base sequence to RNA transcribed from gene, only with U in place of T.

c) A terminator- terminator is present at 3' end of coding strand and defines the end of the process of transcription. The base sequence of the mRNA molecule is complementary to that ofthe antisense strand which served as it template. Like DNA synthesis RNA synthesis also proceeds from 5' to 3' direction (5'-3').

Transcription of mRNA in Eukaryotes:

Eukaryotes-total 4 types of RNA polymerase, 3 types of RNA polymerase in nucleus, one inorganelles,

- **RNA-Polymerase I:** transcribes **rRNA (28S, 18S & 5.8S)**
- **RNA polymerase II:** transcribes **precursor of mRNA (hnRNA**- heterogeneous nuclear RNA)
- **RNA polymerase III:** transcribes **tRNA, 5SrRNA & snRNAs** (smallnuclear RNAs)

1. **Initiation**: binding of **RNA polymerase** to the **promoter region** with the help of an **Initiation Factor- Sigma factor** (binding of σ-factor alter the property of enzyme; make to function as an initiation enzyme).
2. **Elongation-** RNA polymerase will keep on making a complementary strand against template strand with the help of ribonucleotides. The newly transcribed strand keeps separating and the DNA duplex keep on folding back instantaneously. During elongation, same RNA polymerase acts as elongation enzyme due to separation of σ- factor from it. **The direction of transcription is also from 5' 3'like replication**. So the template against which it is transcribed has polarity of 3'—5'.
3. **Termination**- after reaching the terminator region newly formed or nascent RNA falls off along with RNA polymerase. Termination is assisted by Rho-factor(ρ-factor)

   In eukaryotes the promoter site is recognized by presence of specific nucleotide sequence called **TATA box or Hogness box or Pribnow Box** (7 base pair long-TATAAA or TATAATs) located 19-27bp upstream to the start point. Another sequence is CAAT box present between -70 and -80bp. The nucleotide sequence at the two ends of all mRNA molecules is the same. Normally mRNA carries the

codons of signal complete protein molecule (monocistronic mRNA) in eukaryotes, but in prokaryotes, it carries codons from several adjacent DNA cistron and becomes much longer in size (polycistronic mRNA).

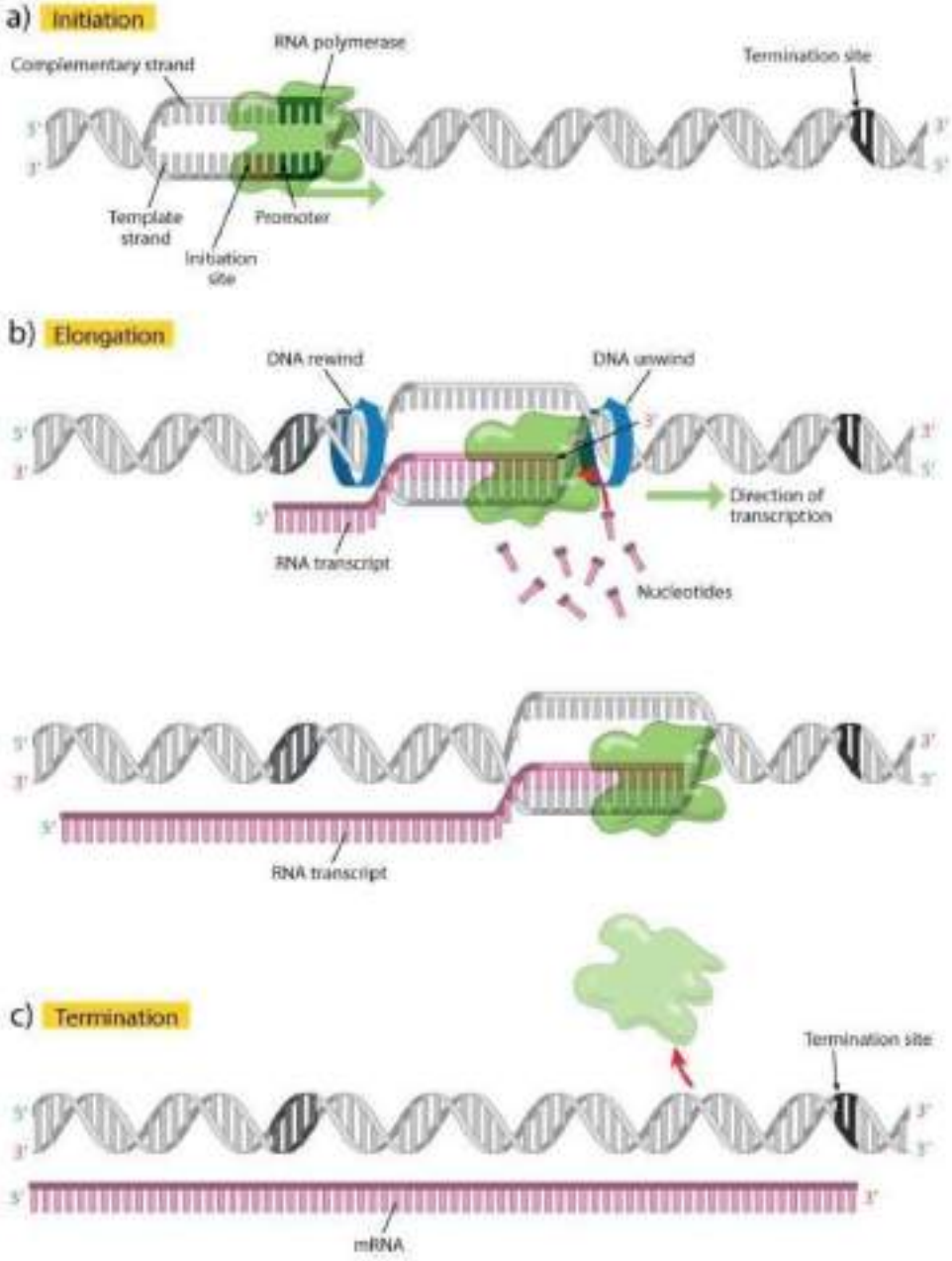


Figure 33: Eukaryotic cells transcription process.

Processing of Eukaryotic Transcript:

- **Splicing-** removal of **non-functional introns** and joining of all **functional exons** to make it a functional transcript. Splicing is important to remove the non-functional partof genetic information the DNA has kept but RNA does not need it. During copying from DNA, RNA does receive this non informative part in the form of introns but remove it with the help of some enzymes to make it functional.
- **Capping-** addition of methyl-guanosine triphosphate at 5' end of hnRNA
- **Tailing-** addition of 200-300 adenylated nucleotides at 3'end of hnRNA, addition of these nucleotides has no relation with the template.

The fully processed hnRNA is called mRNA, transported to the cytoplasm for translation.

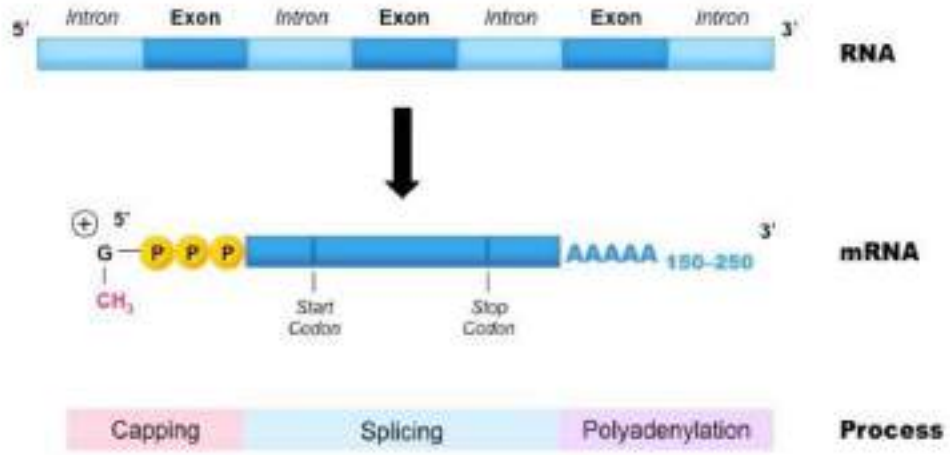


Figure 34: Processing of Eukaryotic Transcript

II- Translation process:

Components of Translation:

- **mRNA**– the mRNA serves at the template that will determine the sequence of aminoacids in the new polypeptide. It has following components:
  - 5' untranslated region or 5'UTR.
  - Initiation codon.
  - Coding region.
  - Stop codon.
  - 3' untranslated region or 3'UTR.
- **t-RNA-** tRNA, a clover leaf shaped molecule, delivers the correct amino acid to the ribosome as directed by the codon on the mRNA for incorporation into the polypeptide. It has following arms, each with specified function:
  - 3'amino acid carrier arm or acceptor arm with –CCA sequence.
  - Ribosome recognizing arm-to recognize A or P or E-site.
  - Anticodon arm- with 3 nucleotides to bind to complementary codon.
  - Enzyme recognizing arm- to recognize specific aminoacyl synthetase.
  - 5énd with G.
- **Ribosome-** protein synthesizing machinery, help in holding mRNA and tRNA forspecific codon translation, has following components:
  - Smaller subunit (30S or 40S).
  - Larger subunit (50S or 60S).
  - Groove or tunnel between two subunits to hold mRNA.
  - Three sites- P, A and E-site.
  - Enzyme, peptidyl transferase, helps in peptide bond formation.

General steps of eukaryote translation:

1- **Activation of amino acids or charging of amino acids:** the amino acids attachment to the tRNA molecules is an active process and requires a lot of energy. In the presence of ATP, an amino acid combines with its specific amino acyl-tRNA synthetase; Mg2+ is also required in this reaction.

2- **Aminoacylation of tRNA or charging of tRNA:** It is the loading of tRNA with the activated amino acid.

3- **Initiation of translation**: In the first step there is binding of mRNA with smaller subunit of ribosome. Translation of Initiation codon (AUG) by a charged tRNA with Methionine (n-formyl methionine, f-Met, in prokaryote) amino acids takes place. It is followed by the translation of second codon by 2nd charged tRNA. After the translation of first two codons, the association of bigger subunit of ribosome takes place to form a complete translational complex. When two such charged tRNA comes close, the peptide bond between two amino acids, they carry, will take place with the help of a ribozyme called- Peptidyl transferase (23SrRNA molecule) enzyme. Formation of peptide bond between 1st& 2nd amino acid takes place. UTR- (Un- Translated-Regions) is the flanks of mRNA before Initiation and after the stop codon, which are not to be translated, but they play role in efficient translation.

4- **Elongation:** The translated part of mRNA translocate from one to next codon. Regular addition of new amino acids takes place at A Polypeptide chain (PPC) keeps elongating at the expense of energy provided by GTP. PPC hangs in the groove of bigger subunit of ribosome on the P-site.

5- **Termination-** Binding of releasing factors to the stop codon helps in the release of polypeptide and terminates translation. Synthesis of polypeptide terminates

when a nonsense codon of mRNA reaches the A-site. There are three nonsense codons- UAA, UAG & UGA. These codons are not recognized by any of the tRNAs. There is no tRNA having anticodon complementary to stop codon i.e., none of the tRNA has AUU, AUC or ACU anticodon. Finally, the ribosome encounters a stop codon. The polypeptide, tRNA and mRNA are released. The small and large subunits dissociate from one another.

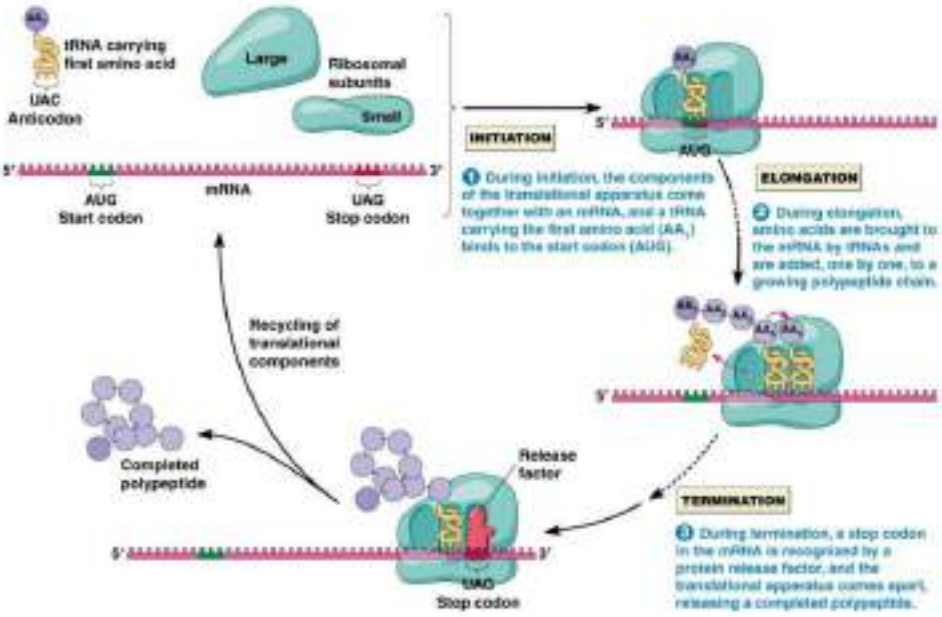


Figure 35: Eukaryote translation process.

GENE REGULATION

I- Prokaryotic Gene Regulation

The DNA of prokaryotes is organized into a circular chromosome, supercoiled within the nucleoid region of the cell cytoplasm. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are encoded together in blocks called **operons**. For example, all of the genes needed to use lactose as an energy source are coded next to each other in the lactose (or lac) operon and transcribed into a single mRNA.

In prokaryotic cells, there are three types of regulatory molecules that can affect the expression of operons: repressors, activators, and inducers. Repressors and activators are proteins produced in the cell. Both repressors and activators regulate gene expression by binding to specific DNA sites adjacent to the genes they control. In general, activators bind to the promoter site, while repressors bind to operator regions. **Repressors** prevent transcription of a gene in response to an external stimulus, whereas **activators** increase the transcription of a gene in response to an external stimulus. Inducers are small molecules that may be produced by the cell or that are in the cell's environment. Inducers either activate or repress transcription depending on the needs of the cell and the availability of substrate.

The operon:

According to the operon model, several gene codes for an enzyme in some metabolic pathways are located in sequence on chromosome. The expressions of structural genes are controlled by some regulatory genes. The **Operon means a unit of gene expression** and regulation which typically includes:

1- **The structural genes:** also called cistron are any gene/s other than the regulatory genes, whose products or enzymes are involved in a specific biosynthetic pathway and whose expression is coordinately controlled.

2- **Operator sequence**: control elements such as an operator sequence, which is a DNA sequence that regulates transcription of the structural genes.

3- **Regulator gene (s):** the genes, whose products recognize the control elements e.g., a repressor which binds told regulates the operator sequence of the same operon.

Operon has structural and regulatory genes that function as a single unit, it includes the following:

- A regulator gene is located outside the operon codes for a repressor or Apo-repressor protein molecule.
- A promoter is a sequence of DNA where RNA polymerase attaches when a gene is to be transcribed.
- An operator is a short sequence of DNA where repressor binds, preventing RNA polymerase from attaching to the promoter.
- Structural genes code for enzymes of a metabolic pathway and are transcribed as a unit.

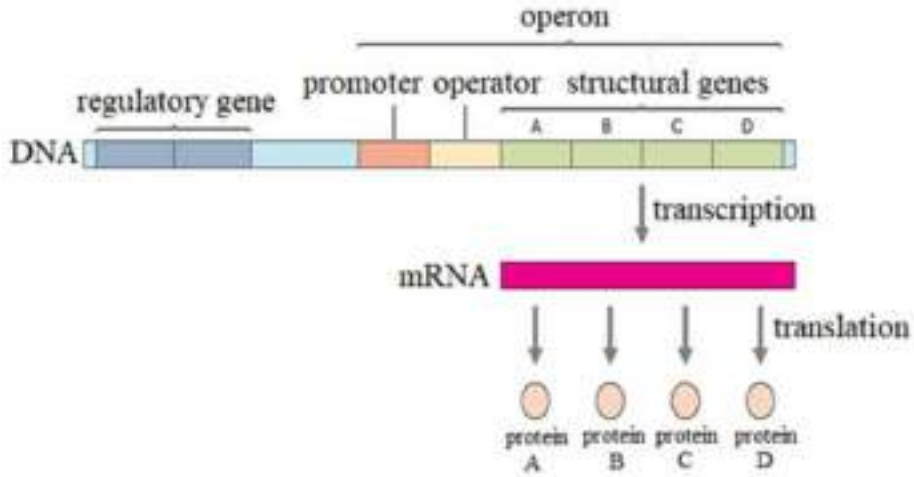


Figure 36: The operon structure.

The trp Operon: A Repressible Operon

Bacteria such as Escherichia coli need amino acids to survive, and are able to synthesize many of them. **Tryptophan** is one such amino acid that E. coli can either ingest from the environment or synthesize using enzymes that are encoded by five genes. These five genes are next to each other in what is called the **tryptophan (trp) operon** (Figure 37). The genes are transcribed into a single mRNA, which is then translated to produce all five enzymes. If tryptophan is present in the environment, then E. coli does not need to synthesize it and the trp operon is switched off. However, when tryptophan availability is low, the switch controlling the operon is turned on, the mRNA is transcribed, the enzyme proteins are translated, and tryptophan is synthesized.

The trp operon includes three important regions: the coding region, the trp operator and the trp promoter. The coding region includes the genes for the five tryptophan biosynthesis enzymes. Just before the coding region is the **transcriptional start site**. The promoter sequence, to which RNA polymerase binds to initiate transcription, is before or "upstream" of the transcriptional start site. Between the promoter and the transcriptional start site is the operator region.

The trp **operator** contains the DNA code to which the trp repressor protein can bind. However, the repressor alone cannot bind to the operator. When tryptophan is present in the cell, two tryptophan molecules bind to the trp repressor, which changes the shape of the repressor protein to a form that can bind to the trp operator. Binding of the tryptophan–repressor complex at the operator physically prevents the RNA polymerase from binding to the promoter and transcribing the downstream genes.

When tryptophan is not present in the cell, the repressor by itself does not bind to the operator, the polymerase can transcribe the enzyme genes, and tryptophan is synthesized. Because the repressor protein actively binds to the operator to keep the

genes turned off, the trp operon is said to be negatively regulated and the proteins that bind to the operator to silence trp expression are **negative regulators**

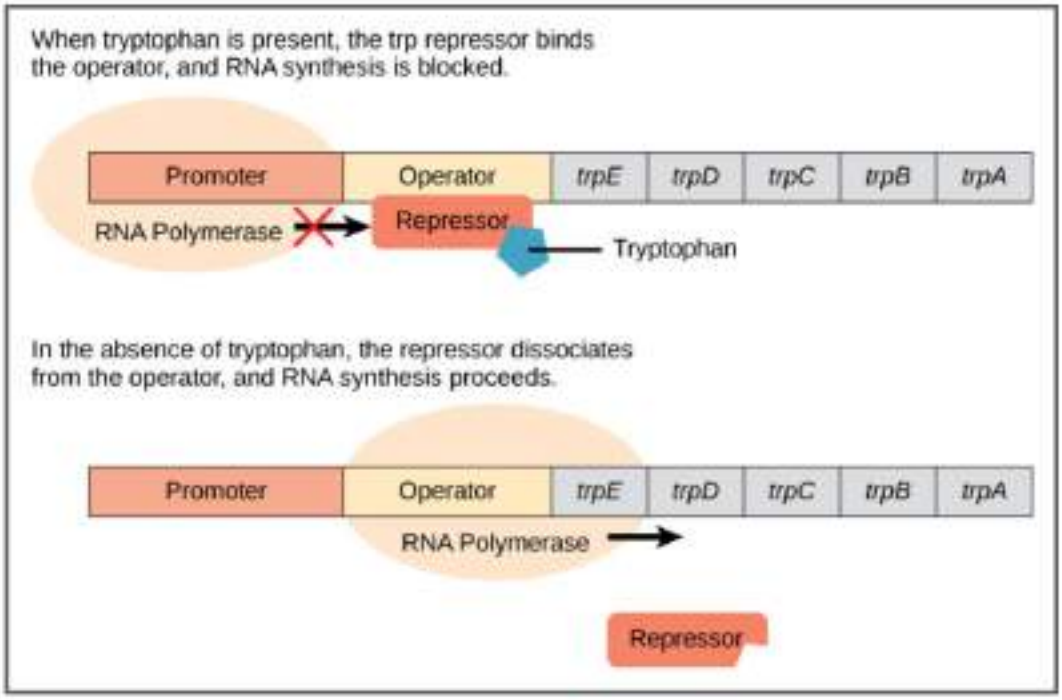


Figure 37: The tryptophan operon. The five genes that are needed to synthesize tryptophan in E. coli are located next to each other in the trp operon. When tryptophan is plentiful, two tryptophan molecules bind the repressor protein at the operator sequence. This physically blocks the RNA polymerase from transcribing the tryptophan genes. When tryptophan is absent, the repressor protein does not bind to the operator and the genes are transcribed

Catabolite Activator Protein (CAP): A Transcriptional Activator

Just as the trp operon is negatively regulated by tryptophan molecules, there are proteins that bind to the promoter sequences that act as **positive regulators** to turn genes on and activate them. For example, when glucose is scarce, E. coli bacteria can turn to other sugar sources for fuel. To do this, new genes to process these alternate sugars must be transcribed. When glucose levels drop, cyclic AMP (cAMP) begins to accumulate in the cell. The cAMP molecule is a signaling molecule that is involved in glucose and energy metabolism in E. coli. Accumulating cAMP binds to

the positive regulator **catabolite activator protein (CAP)**, a protein that binds to the promoters of operons which control the processing of alternative sugars. When cAMP binds to CAP, the complex then binds to the promoter region of the genes that are needed to use the alternate sugar sources (Figure 38). In these operons, a CAP-binding site is located upstream of the RNA-polymerase-binding site in the promoter. CAP binding stabilizes the binding of RNA polymerase to the promoter region and increases transcription of the associated protein-coding genes.

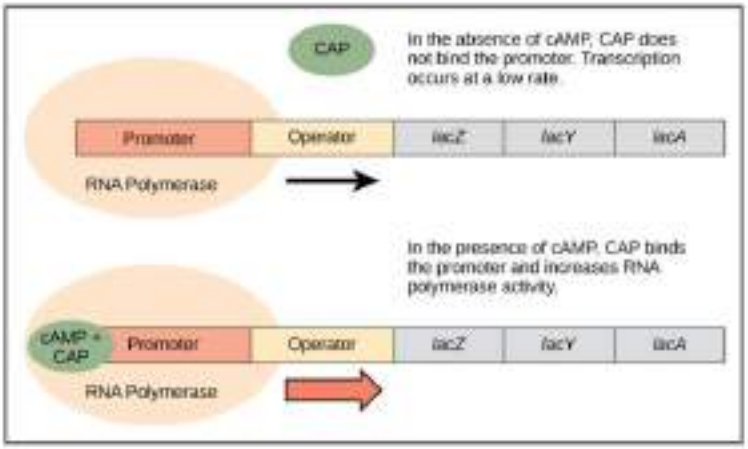


Figure 38: Transcriptional activation by the CAP protein. When glucose levels fall, E. coli may use other sugars for fuel but must transcribe new genes to do so. As glucose supplies become limited, cAMP levels increase. This cAMP binds to the CAP protein, a positive regulator that binds to a promoter region upstream of the genes required to use other sugar sources.

The lac Operon: An Inducible Operon

The third type of gene regulation in prokaryotic cells occurs through inducible operons, which have proteins that bind to activate or repress transcription depending on the local environment and the needs of the cell. The lac operon is a typical inducible operon. As mentioned previously, E. coli is able to use other sugars as energy sources when glucose concentrations are low.

The **lac operon** encodes the genes necessary to acquire and process the lactose from the local environment. The Z gene of the lac operon encodes beta-galactosidase, which breaks lactose down to glucose and galactose.

However, for the lac operon to be activated, two conditions must be met. First, the level of glucose must be very low or nonexistent. Second, lactose must be present. Only when glucose is absent and lactose is present will the lac operon be transcribed (Figure 39). In the absence of glucose, the binding of the CAP protein makes transcription of the lac operon more effective. When lactose is present, its metabolite, allolactose, binds to the lac repressor and changes its shape so that it cannot bind to the lac operator to prevent transcription. This combination of conditions makes sense for the cell, because it would be energetically wasteful to synthesize the enzymes to process lactose if glucose was plentiful or lactose was not available. It should be mentioned that the lac operon is transcribed at a very low rate even when glucose is present and lactose absent.

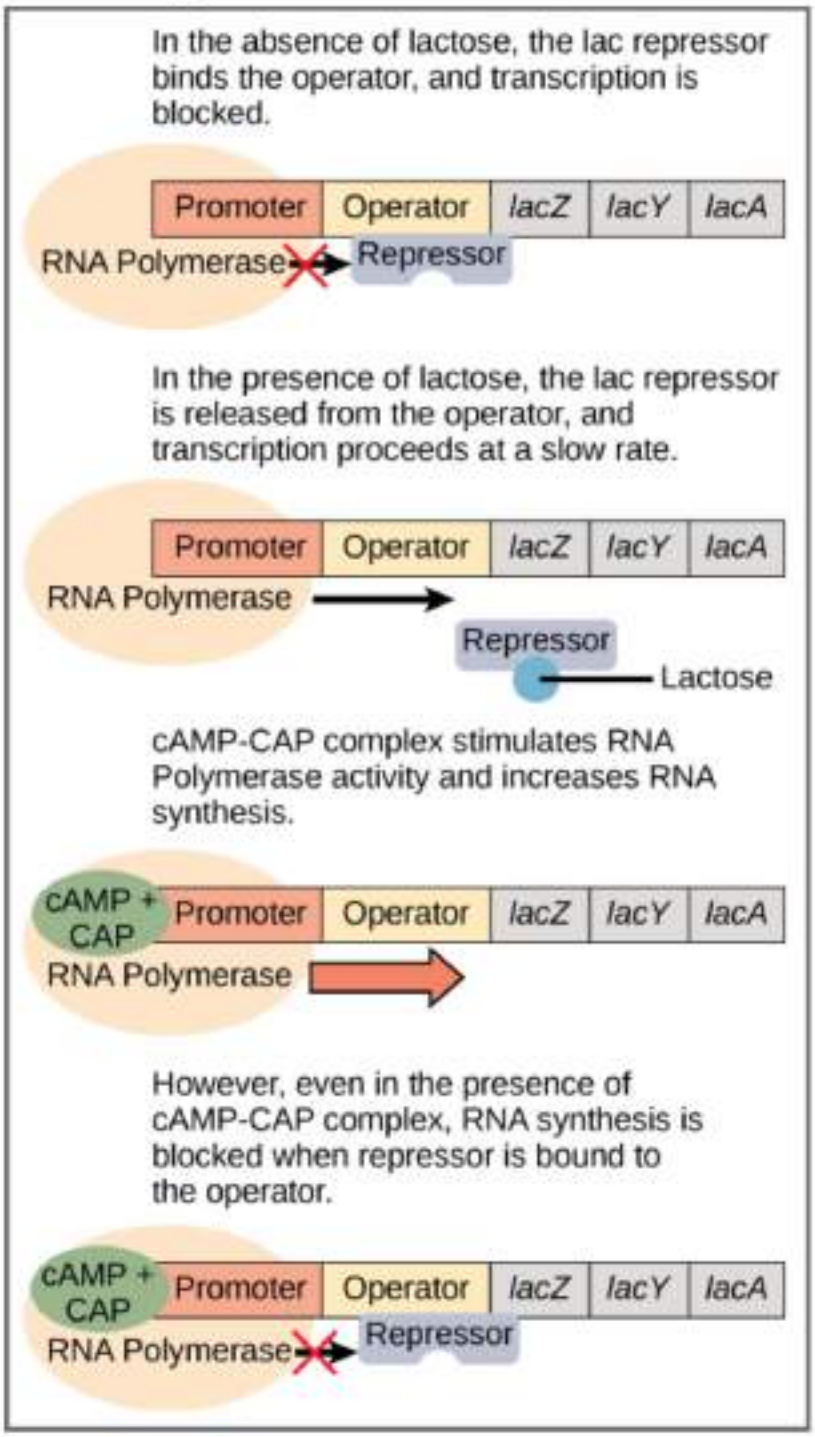


Figure 39: Regulation of the lac operon. Transcription of the lac operon is carefully regulated so that its expression only occurs when glucose is limited and lactose is present to serve as an alternative fuel source.

Eukaryotic Epigenetic Gene Regulation:

Eukaryotic gene expression is more complex than prokaryotic gene expression because the processes of transcription and translation are physically separated. Unlike prokaryotic cells, eukaryotic cells can regulate gene expression at many different levels. Epigenetic changes are inheritable changes in gene expression that do not result from changes in the DNA sequence. Eukaryotic gene expression begins with control of access to the DNA. Transcriptional access to the DNA can be controlled in two general ways: chromatin remodeling and DNA methylation. Chromatin remodeling changes the way that DNA is associated with chromosomal histones. DNA methylation is associated with developmental changes and gene silencing.

Epigenetic Control: Regulating Access to Genes within the Chromosome:

The human genome encodes over 20,000 genes, with hundreds to thousands of genes on each of the 23 human chromosomes. The DNA in the nucleus is precisely wound, folded, and compacted into chromosomes so that it will fit into the nucleus. It is also organized so that specific segments can be accessed as needed by a specific cell type. The first level of organization, or **packing**, is the winding of DNA strands around histone proteins. Histones package and order DNA into structural units called nucleosome complexes, which can control the access of proteins to the DNA regions (Figure 40a). Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string (Figure 40b). These beads (histone proteins) can move along the string (DNA) to expose different sections of the molecule. If DNA encoding a specific gene is to be transcribed into RNA, the nucleosomes surrounding that region of DNA can slide down the DNA to open that specific chromosomal region and allow for the transcriptional machinery (RNA polymerase) to initiate transcription (Figure 41).

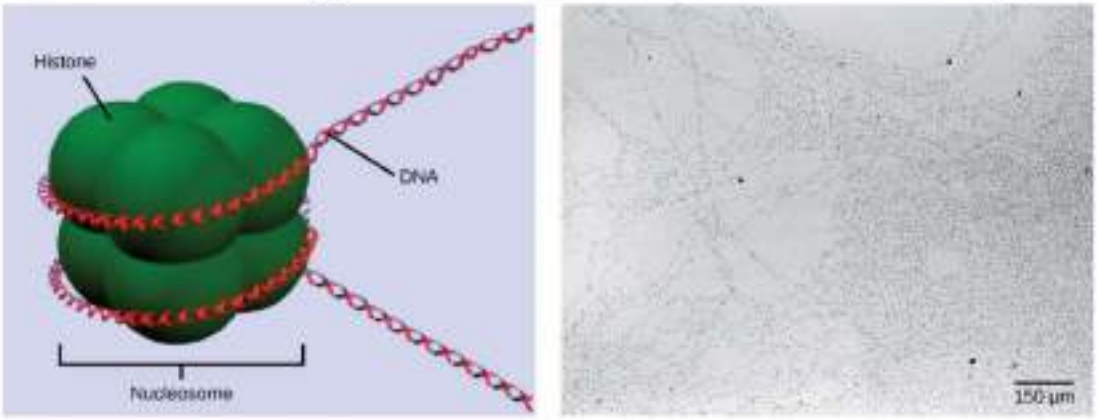


Figure 40: DNA is folded around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit "micrograph": modification of work by Chris Woodcock)

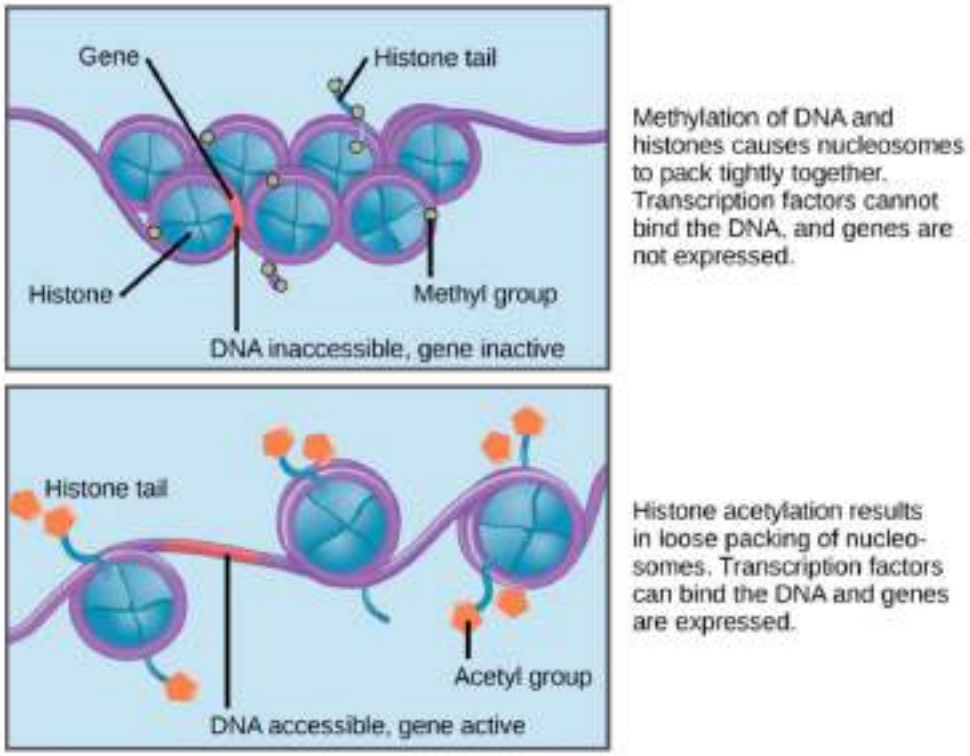


Figure 41: Nucleosomes can slide along DNA. When nucleosomes are spaced closely together (top), transcription factors cannot bind and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed. Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

In females, one of the two X chromosomes is inactivated during embryonic development because of epigenetic changes to the chromatin.

How closely the histone proteins associate with the DNA is regulated by signals found on both the histone proteins and on the DNA. These signals are functional groups added to histone proteins or to DNA and determine whether a chromosomal region should be open or closed (Figure 42 depicts modifications to histone proteins and DNA). These tags are not permanent but may be added or removed as needed. Some chemical groups (phosphate, methyl, or acetyl groups) are attached to specific amino acids in histone "tails" at the N-terminus of the protein. These groups do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins. DNA is a negatively charged molecule and unmodified histones are positively charged; therefore, changes in the charge of the histone will change how tightly wound the DNA molecule will be. By adding chemical modifications like acetyl groups, the charge becomes less positive, and the binding of DNA to the histones is relaxed. Altering the location of nucleosomes and the tightness of histone binding opens some regions of chromatin to transcription and closes others.

The DNA molecule itself can also be modified by methylation. DNA methylation occurs within very specific regions called CpG islands. These are stretches with a high frequency of cytosine and guanine dinucleotide DNA pairs (CG) found in the promoter regions of genes. The cytosine member of the CG pair can be methylated (a methyl group is added). Methylated genes are usually silenced, although methylation may have other regulatory effects. In some cases, genes that are silenced during the development of the gametes of one parent are transmitted in their silenced condition to the offspring. Such genes are said to be imprinted. Parental diet or other environmental conditions may also affect the methylation patterns of genes, which in turn modifies gene expression. Changes in chromatin organization interact with

DNA methylation. DNA methyltransferases appear to be attracted to chromatin regions with specific histone modifications. Highly methylated (hypermethylated) DNA regions with deacetylated histones are tightly coiled and transcriptionally inactive.

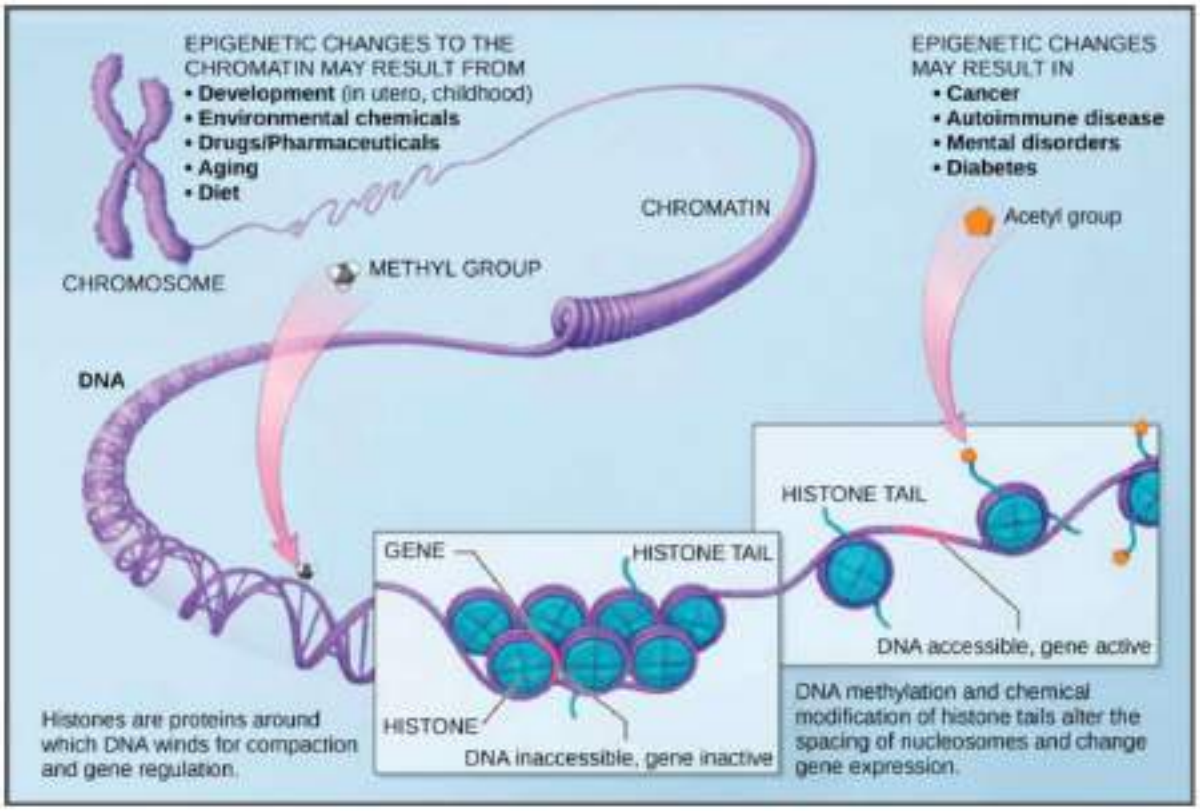


Figure 42: Histone proteins and DNA nucleotides can be modified chemically. Modifications affect nucleosome spacing and gene expression. (credit: modification of work by NIH)

Epigenetic changes are not permanent, although they often persist through multiple rounds of cell division and may even cross generational lines. Chromatin remodeling alters the chromosomal structure (open or closed) as needed. If a gene is to be transcribed, the histone proteins and DNA in the chromosomal region encoding that gene are modified in a way that opens the promoter region to allow RNA polymerase and other proteins, called **transcription factors**, to bind and initiate transcription. If a gene is to remain turned off, or silenced, the histone proteins and DNA have

different modifications that signal a closed chromosomal configuration. In this closed configuration, the RNA polymerase and transcription factors do not have access to the DNA and transcription cannot occur (Figure 42).

II- Eukaryotic Transcription Gene Regulation

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

The Promoter and the Transcription Machinery

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

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

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

Enhancers and Transcription

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

Enhancer regions are binding sequences, or sites, for specific transcription factors. When a protein transcription factor binds to its enhancer sequence, the shape of the protein changes, allowing it to interact with proteins at the promoter site. However, since the enhancer region may be distant from the promoter, the DNA must bend to allow the proteins at the two sites to come into contact. DNA bending proteins help to bend the DNA and bring the enhancer and promoter regions together (Figure 43). This shape change allows for the interaction of the specific activator proteins bound to the enhancers with the general transcription factors bound to the promoter region and the RNA polymerase.

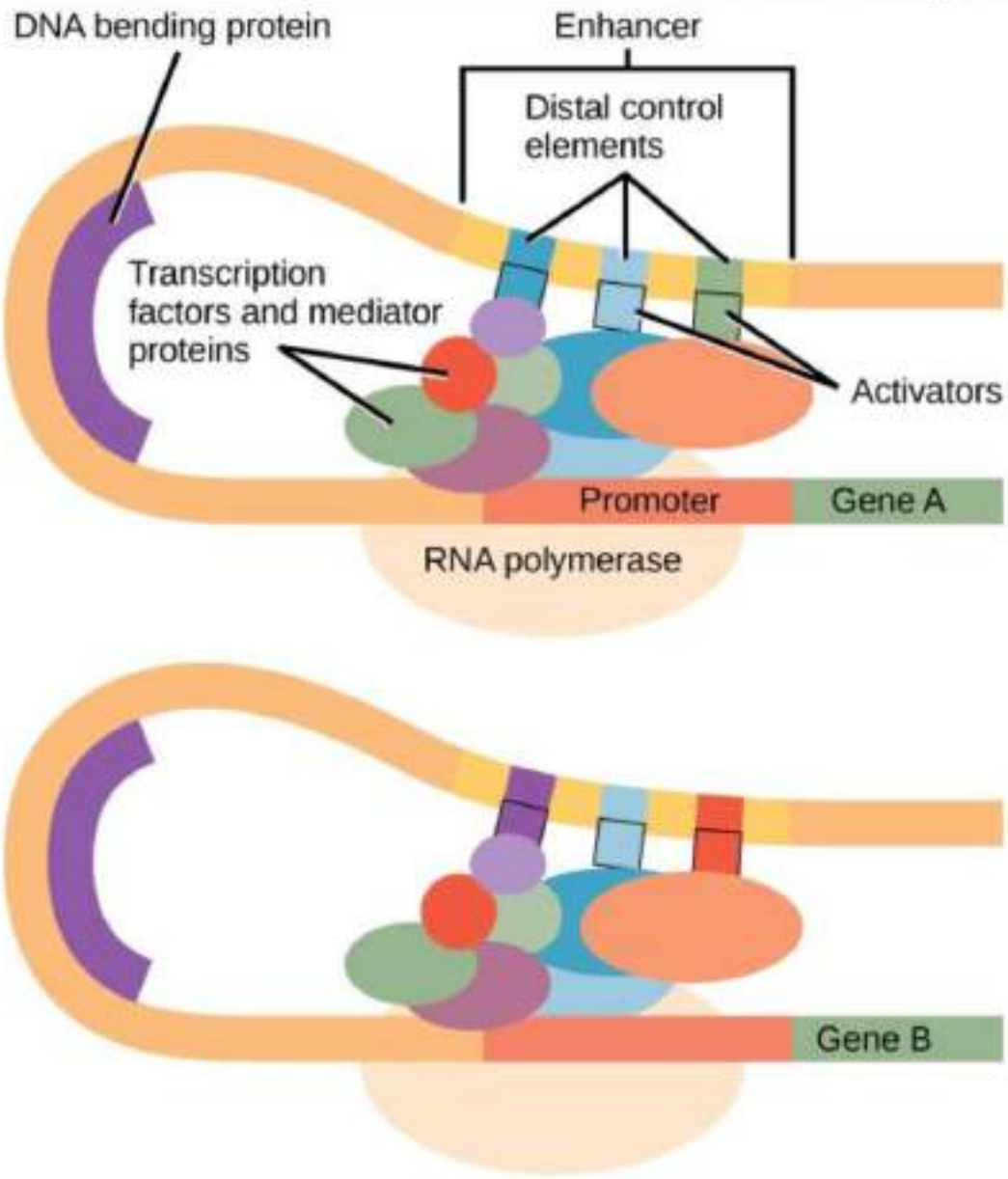


Figure 43: Interaction between proteins at the promoter and enhancer sites. An enhancer is a DNA sequence that promotes transcription. Each enhancer is made up of short DNA sequences called distal control elements. Activators bound to the distal control elements interact with mediator proteins and transcription factors. Two different genes may have the same promoter but different distal control elements, enabling differential gene expression.

Turning Genes Off: Transcriptional Repressors

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

Eukaryotic Post-transcriptional Gene Regulation

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

RNA Splicing, the First Stage of Post-transcriptional Control

In eukaryotic cells, the RNA transcript often contains regions, called introns, that are removed prior to translation. The regions of RNA that code for protein are called **exons**. (Figure 44). After an RNA molecule has been transcribed, but prior to its departure from the nucleus to be translated, the RNA is processed and the introns are removed by splicing. Splicing is done by spliceosomes, ribonucleoprotein complexes that can recognize the two ends of the intron, cut the transcript at those two points, and bring the exons together for ligation.

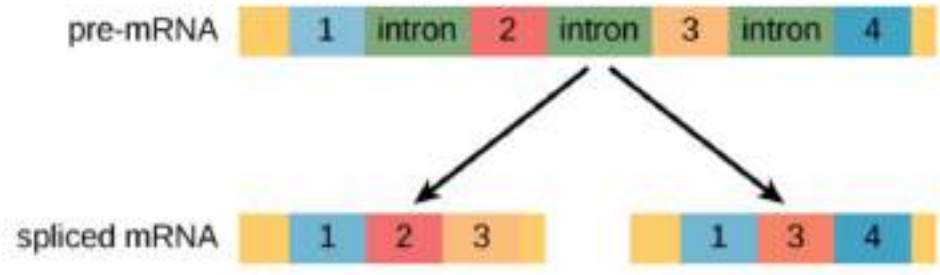


Figure 44: Pre-mRNA can be alternatively spliced to create different proteins

Alternative RNA Splicing

In the 1970s, genes were first observed that exhibited alternative RNA splicing. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different combinations of exons are combined to form the mRNA (Figure 45). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of gene regulation, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells or at different stages of development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70 percent of genes in humans are expressed as multiple proteins through alternative splicing. Although there are multiple ways to alternatively splice RNA transcripts, the original 5'-3' order of the exons is always conserved. That is, a transcript with exons 1 2 3 4 5 6 7 might be spliced 1 2 4 5 6 7 or 1 2 3 6 7, but never 1 2 5 4 3 6 7.

How could alternative splicing evolve? Introns have a beginning- and ending-recognition sequence; it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and instead find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such intron skipping, but mutations are likely to lead to their failure. Such "mistakes" would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is abnormal splicing rather than mutations in a coding sequence. However, alternative splicing could possibly create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way by providing genes that may evolve without eliminating the original, functional protein.

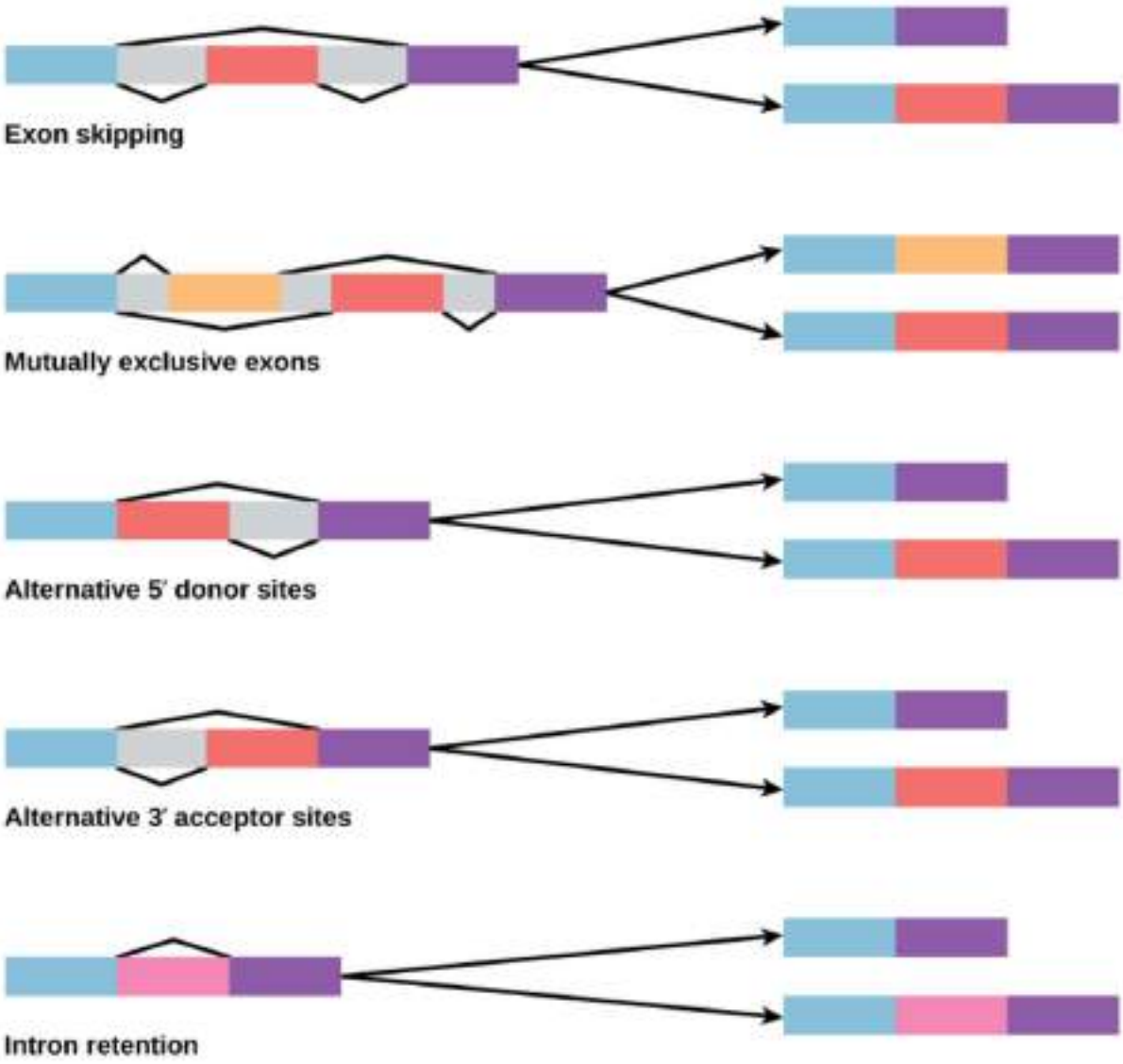


Figure 45: There are five basic modes of alternative splicing.

Control of RNA Stability

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

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

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

region is called the **3' UTR** (Figure 46). The binding of RBPs to these regions can increase or decrease the stability of an RNA molecule, depending on the specific RBP that binds.

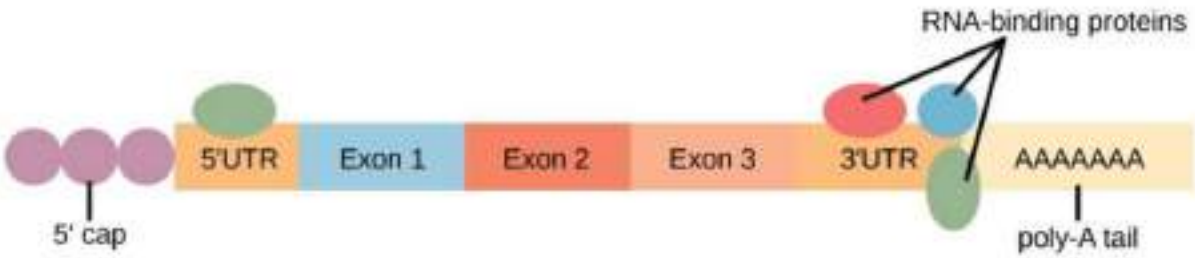


Figure 46: RNA-binding proteins. The protein-coding region of this processed mRNA is flanked by 5' and 3' untranslated regions (UTRs). The presence of RNA-binding proteins at the 5' or 3' UTR influences the stability of the RNA molecule.

RNA Stability and microRNAs

In addition to RBPs that bind to and control (increase or decrease) RNA stability, other elements called microRNAs can bind to the RNA molecule. These **microRNAs**, or miRNAs, are short RNA molecules that are only 21 to 24 nucleotides in length. The miRNAs are made in the nucleus as longer pre-miRNAs. These pre-miRNAs are chopped into mature miRNAs by a protein called **Dicer**. Like transcription factors and RBPs, mature miRNAs recognize a specific sequence and bind to the RNA; however, miRNAs also associate with a ribonucleoprotein complex called the **RNA-induced silencing complex (RISC)**. The RNA component of the RISC base-pairs with complementary sequences on an mRNA and either impede translation of the message or lead to the degradation of the mRNA.

Eukaryotic Translational and Post-translational Gene Regulation

After RNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. As previously

discussed, the stability of the RNA will have a large impact on its translation into a protein. As the stability changes, the amount of time that it is available for translation also changes.

The Initiation Complex and Translation Rate

Like transcription, translation is controlled by proteins that bind and initiate the process. In translation, the complex that assembles to start the process is referred to as the translation **initiation complex**. In eukaryotes, translation is initiated by binding the initiating met-tRNAi to the 40S ribosome. This tRNA is brought to the 40S ribosome by a protein initiation factor, **eukaryotic initiation factor-2 (eIF-2)**. The eIF-2 protein binds to the high-energy molecule **guanosine triphosphate (GTP)**. The tRNA-eIF2-GTP complex then binds to the 40S ribosome. A second complex forms on the mRNA. Several different initiation factors recognize the 5' cap of the mRNA and proteins bound to the poly-A tail of the same mRNA, forming the mRNA into a loop. The cap-binding protein eIF4F brings the mRNA complex together with the 40S ribosome complex. The ribosome then scans along the mRNA until it finds a start codon AUG. When the anticodon of the initiator tRNA and the start codon are aligned, the GTP is hydrolyzed, the initiation factors are released, and the large **60S ribosomal subunit** binds to form the translation complex. The binding of eIF-2 to the RNA is controlled by phosphorylation. If eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP. Therefore, the initiation complex cannot form properly and translation is impeded (Figure 47). When eIF-2 remains unphosphorylated, the initiation complex can form normally, and translation can proceed.

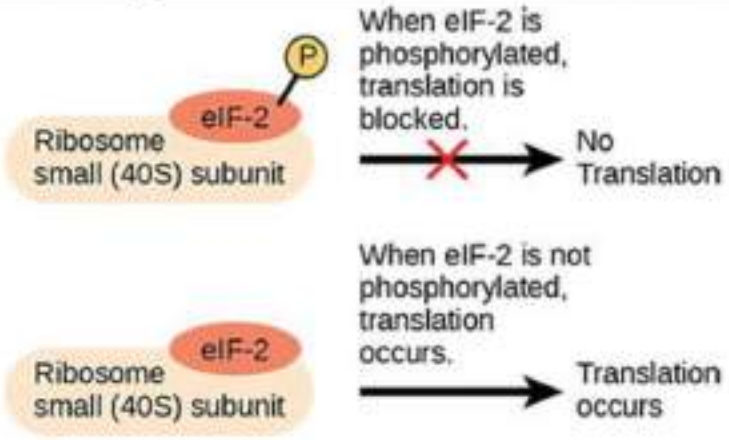


Figure 47: Gene expression can be controlled by factors that bind the translation initiation complex

Chemical Modifications, Protein Activity, and Longevity

Proteins can be chemically modified with the addition of groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins regulates their activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, in the cytoplasm, or attached to the plasma membrane. Chemical modifications occur in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. These changes can alter epigenetic accessibility, transcription, mRNA stability, or translation—all resulting in changes in expression of various genes. This is an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment. Because proteins are involved in every stage of gene regulation, the phosphorylation of a protein (depending on the protein that is modified) can alter accessibility to the chromosome, can alter translation (by altering transcription factor binding or function), can change nuclear shuttling (by influencing modifications to the nuclear pore complex), can alter RNA stability (by binding or not binding to the RNA to regulate its stability), can modify translation (increase or decrease), or can change

post-translational modifications (add or remove phosphates or other chemical modifications).

The addition of an ubiquitin group to a protein marks that protein for degradation. Ubiquitin acts like a flag indicating that the protein lifespan is complete. These proteins are moved to the **proteasome**, an organelle that functions to remove proteins, to be degraded (Figure 48). One way to control gene expression, therefore, is to alter the longevity of the protein.

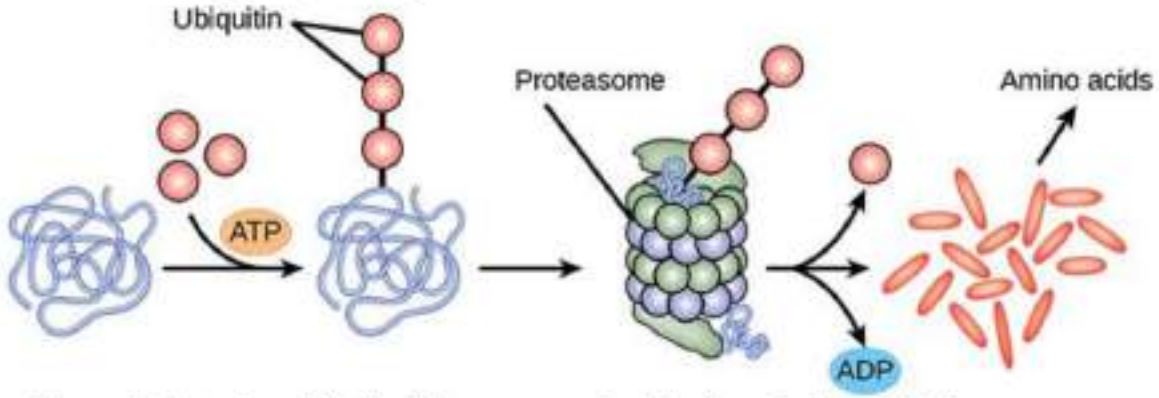


Figure 48: Proteins with ubiquitin tags are marked for degradation within the proteasome

Cancer and Gene Regulation

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

For cells to move through each phase of the cell cycle, the cell must pass through checkpoints. This ensures that the cell has properly completed the step and has not encountered any mutation that will alter its function. Many proteins, including cyclin B, control these checkpoints. The phosphorylation of cyclin B, a post-translational

event, alters its function. As a result, cells can progress through the cell cycle unimpeded, even if mutations exist in the cell and its growth should be terminated. This posttranslational change of cyclin B prevents it from controlling the cell cycle and contributes to the development of cancer.

Cancer: Disease of Altered Gene Expression

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

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

Tumor Suppressor Genes, Oncogenes, and Cancer

In normal cells, some genes function to prevent excess, inappropriate cell growth. These are tumor-suppressor genes, which are active in normal cells to prevent uncontrolled cell growth. There are many tumor-suppressor genes in cells. The most

studied tumor-suppressor gene is p53, which is mutated in over 50 percent of all cancer types. The p53 protein itself functions as a transcription factor. It can bind to sites in the promoters of genes to initiate transcription. Therefore, the mutation of p53 in cancer will dramatically alter the transcriptional activity of its target genes.

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

Cancer and Epigenetic Alterations

Silencing genes through epigenetic mechanisms is also very common in cancer cells. There are characteristic modifications to histone proteins and DNA that are associated with silenced genes. In cancer cells, the DNA in the promoter region of silenced genes is methylated on cytosine DNA residues in CpG islands. Histone proteins that surround that region lack the acetylation modification that is present when the genes are expressed in normal cells. This combination of DNA methylation and histone deacetylation (epigenetic modifications that lead to gene silencing) is commonly found in cancer. When these modifications occur, the gene present in that chromosomal region is silenced. Increasingly, scientists understand how epigenetic changes are altered in cancer. Because these changes are temporary and can be

reversed—for example, by preventing the action of the histone deacetylase protein that removes acetyl groups, or by DNA methyl transferase enzymes that add methyl groups to cytosines in DNA—it is possible to design new drugs and new therapies to take advantage of the reversible nature of these processes. Indeed, many researchers are testing how a silenced gene can be switched back on in a cancer cell to help re-establish normal growth patterns.

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

Cancer and Transcriptional Control

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

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

many protein kinases that, in turn, activate many transcription factors which control genes involved in cell growth. New drugs that prevent the activation of EGFR have been developed and are used to treat these cancers.

Cancer and Post-transcriptional Control

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

Cancer and Translational/Post-translational Control

There are many examples of how translational or post-translational modifications of proteins arise in cancer. Modifications are found in cancer cells from the increased translation of a protein to changes in protein phosphorylation to alternative splice variants of a protein. An example of how the expression of an alternative form of a protein can have dramatically different outcomes is seen in colon cancer cells. The c-Flip protein, a protein involved in mediating the cell-death pathway, comes in two

forms: long (c-FLIPL) and short (c-FLIPS). Both forms appear to be involved in initiating controlled cell-death mechanisms in normal cells. However, in colon cancer cells, expression of the long form results in increased cell growth instead of cell death. Clearly, the expression of the wrong protein dramatically alters cell function and contributes to the development of cancer.

New Drugs to Combat Cancer: Targeted Therapies

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

BIOTECHNOLOGY AND GENOMICS

INTRODUCTION The study of nucleic acids began with the discovery of DNA, progressed to the study of genes and small fragments, and has now exploded to the field of genomics. Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species. DNA sequencing technology has contributed to advances in genomics. Just as information technology has led to Google maps that enable people to obtain detailed information about locations around the globe, researchers use genomic information to create similar DNA maps of different organisms. These findings have helped anthropologists to better understand human migration and have aided the medical field through mapping human genetic diseases. Genomic information can contribute to scientific understanding in various ways and knowledge in the field is quickly growing.

Biotechnology is the use of biological agents for technological advancement. Biotechnology was used for breeding livestock and crops long before people understood the scientific basis of these techniques. Since the discovery of the structure of DNA in 1953, the biotechnology field has grown rapidly through both academic research and private companies. The primary applications of this technology are in medicine (vaccine and antibiotic production) and agriculture (crop genetic modification in order to increase yields). Biotechnology also has many industrial applications, such as fermentation, treating oil spills, and producing biofuels.

Basic Techniques to Manipulate Genetic Material (DNA and RNA)

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a

nitrogenous base) linked by phosphodiester bonds. The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases. Exposure to high temperatures (DNA denaturation) can separate the two strands and cooling can reanneal them. The DNA polymerase enzyme can replicate the DNA. Unlike DNA, which is located in the eukaryotic cells' nucleus, RNA molecules leave the nucleus. The most common type of RNA that researchers analyze is the messenger RNA (mRNA) because it represents the proteincoding genes that are actively expressed. However, RNA molecules present some other challenges to analysis, as they are often less stable than DNA.

DNA and RNA Extraction

To study or manipulate nucleic acids, one must first isolate or extract the DNA or RNA from the cells. Researchers use various techniques to extract different types of DNA (Figure 49). Most nucleic acid extraction techniques involve steps to break open the cell and use enzymatic reactions to destroy all macromolecules that are not desired (such as unwanted molecule degradation and separation from the DNA sample). A **lysis buffer** (a solution which is mostly a detergent) breaks cells. Note that lysis means "to split". These enzymes break apart lipid molecules in the cell membranes and nuclear membranes. Enzymes such as **proteases** that break down proteins inactivate macromolecules, and **ribonucleases** (RNAses) that break down RNA. Using alcohol precipitates the DNA. Human genomic DNA is usually visible as a gelatinous, white mass. One can store the DNA samples frozen at –80°C for several years.

Scientists perform RNA analysis to study gene expression patterns in cells. RNA is naturally very unstable because RNAses are commonly present in nature and very

difficult to inactivate. Similar to DNA, RNA extraction involves using various buffers and enzymes to inactivate macromolecules and preserve the RNA.

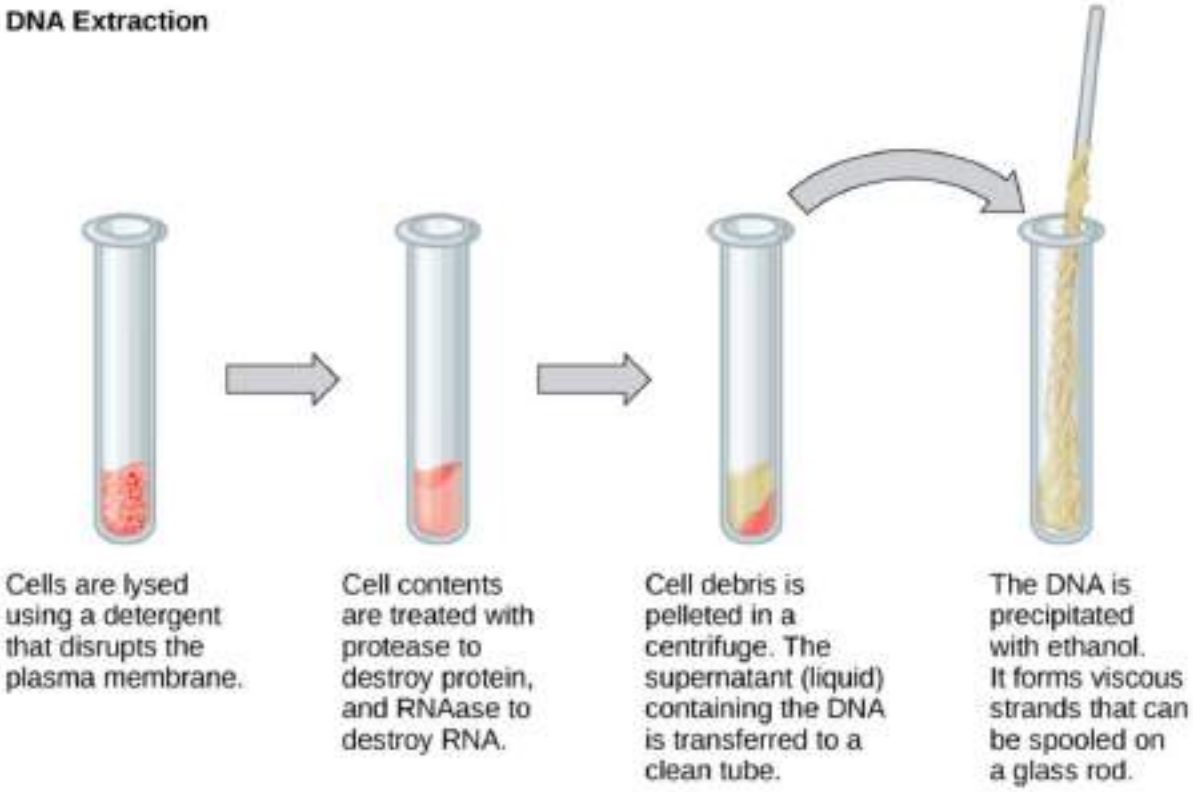
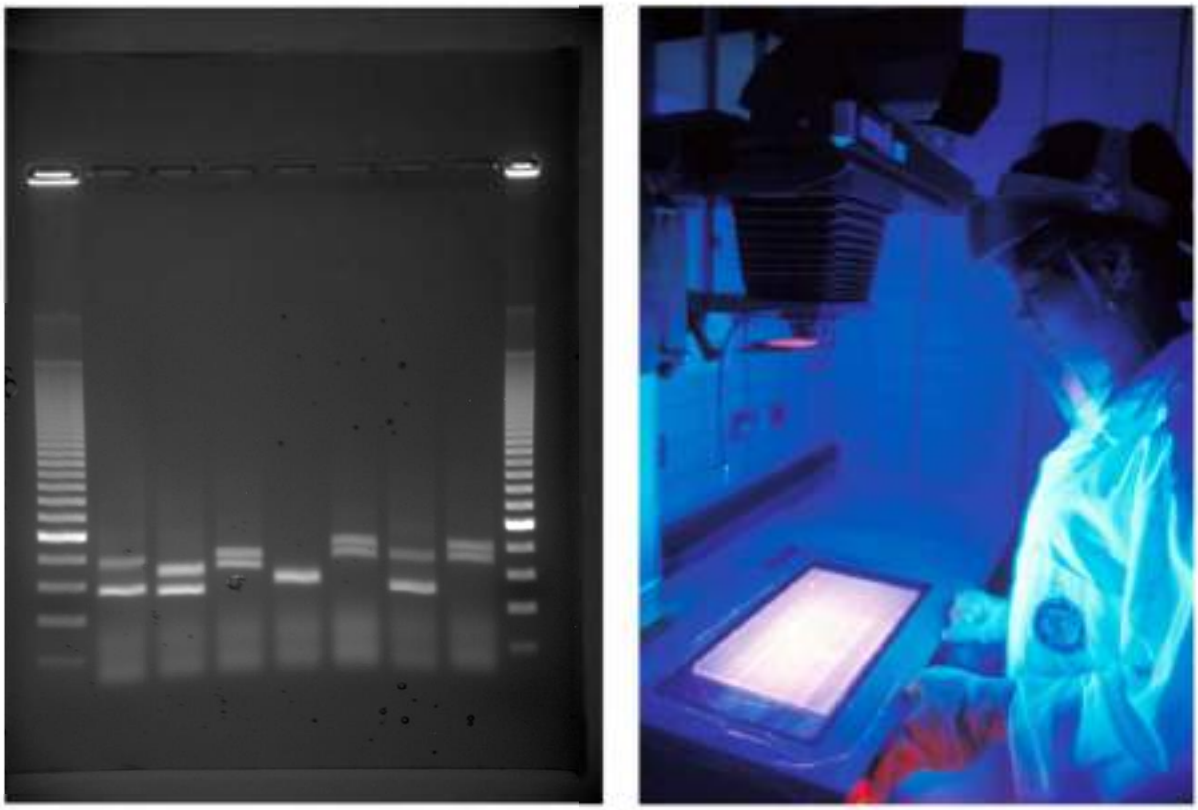


Figure 49: DNA Extraction

Gel Electrophoresis

Because nucleic acids are negatively charged ions at neutral or basic pH in an aqueous environment, an electric field can mobilize them. **Gel electrophoresis** is a technique that scientists use to separate molecules on the basis of size, using this charge. One can separate the nucleic acids as whole chromosomes or fragments. The nucleic acids load into a slot near the semisolid, porous gel matrix's negative electrode, and pulled toward the positive electrode at the gel's opposite end. Smaller molecules move through the gel's pores faster than larger molecules. This difference in the migration rate separates the fragments on the basis of size. There are molecular

weight standard samples that researchers can run alongside the molecules to provide a size comparison. We can observe nucleic acids in a gel matrix using various fluorescent or colored dyes. Distinct nucleic acid fragments appear as bands at specific distances from the gel's top (the negative electrode end) on the basis of their size (Figure 50). A mixture of genomic DNA fragments of varying sizes appear as a long smear; whereas, uncut genomic DNA is usually too large to run through the gel and forms a single large band at the gel's top.



(a) (b)

Figure 50: a) Shown are DNA fragments from seven samples run on a gel, stained with a fluorescent dye, and viewed under UV light; and b) a researcher from International Rice Research Institute, reviewing DNA profiles using UV light. (credit: a: James Jacob, Tompkins Cortland Community College b: International Rice Research Institute)

Nucleic Acid Fragment Amplification by Polymerase Chain Reaction

Although genomic DNA is visible to the naked eye when it is extracted in bulk, DNA analysis often requires focusing on one or more specific genome regions. **Polymerase chain reaction (PCR)** is a technique that scientists use to amplify specific DNA regions for further analysis (Figure 51). Researchers use PCR for many purposes in laboratories, such as cloning gene fragments to analyze genetic diseases, identifying contaminant foreign DNA in a sample, and amplifying DNA for sequencing. More practical applications include determining paternity and detecting genetic diseases.

DNA fragments can also be amplified from an RNA template in a process called **reverse transcriptase PCR (RT-PCR)**. The first step is to recreate the original DNA template strand (called cDNA) by applying DNA nucleotides to the mRNA. This process is called reverse transcription. This requires the presence of an enzyme called reverse transcriptase. After the cDNA is made, regular PCR can be used to amplify it.

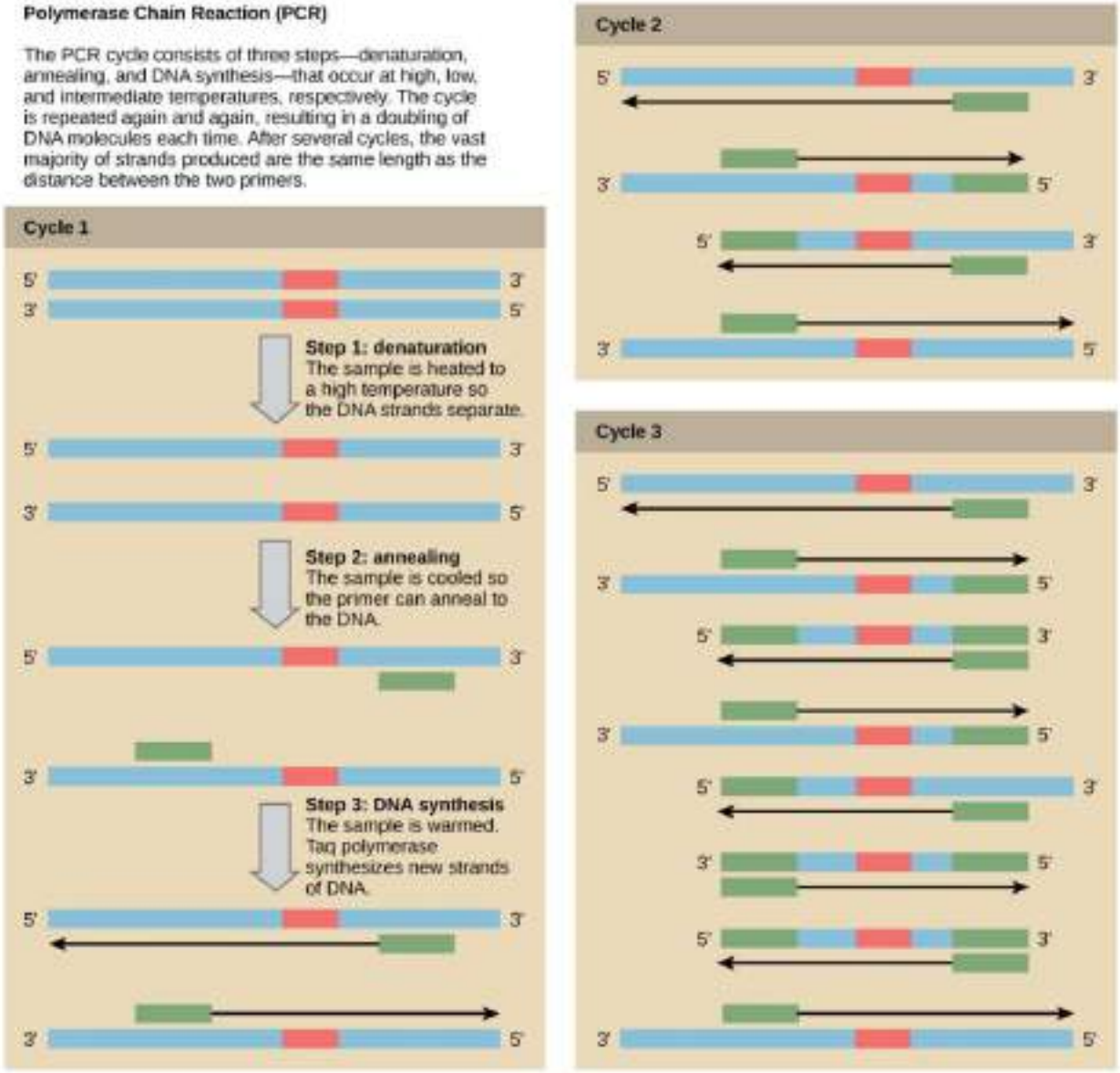


Figure 51: Scientists use polymerase chain reaction, or PCR, to amplify a specific DNA sequence. Primers—short pieces of DNA complementary to each end of the target sequence combine with genomic DNA, Taq polymerase, and deoxynucleotides. Taq polymerase is a DNA polymerase isolated from the thermostable bacterium Thermus aquaticus that is able to withstand the high temperatures that scientists use in PCR. Thermus aquaticus grows in the Lower Geyser Basin of Yellowstone National Park. Reverse transcriptase PCR (RTPCR) is similar to PCR, but cDNA is made from an RNA template before PCR begins

Hybridization, Southern Blotting, and Northern Blotting

Scientists can probe nucleic acid samples, such as fragmented genomic DNA and RNA extracts, for the presence of certain sequences. Scientists design and label short DNA fragments, or **probes** with radioactive or fluorescent dyes to aid detection. Gel electrophoresis separates the nucleic acid fragments according to their size. Scientists then transfer the fragments in the gel onto a nylon membrane in a procedure we call **blotting** (Figure 52). Scientists can then probe the nucleic acid fragments that are bound to the membrane's surface with specific radioactively or fluorescently labeled probe sequences. When scientists transfer DNA to a nylon membrane, they refer to the technique as **Southern blotting**. When they transfer the RNA to a nylon membrane, they call it **Northern blotting**. Scientists use Southern blots to detect the presence of certain DNA sequences in a given genome, and Northern blots to detect gene expression.

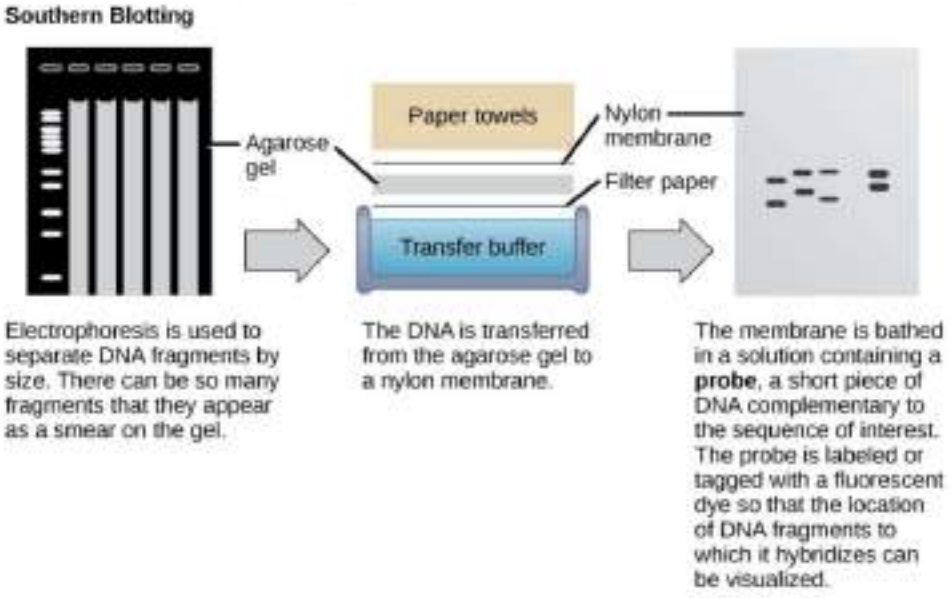


Figure 52: Scientists use Southern blotting to find a particular sequence in a DNA sample. Scientists separate DNA fragments on a gel, transfer them to a nylon membrane, and incubate them with a DNA probe complementary to the sequence of interest. Northern blotting is similar to Southern blotting, but scientists run RNA on the gel instead of DNA. In Western blotting, scientists run proteins on a gel and detect them using antibodies.

Molecular Cloning

In general, the word "cloning" means the creation of a perfect replica; however, in biology, the re-creation of a whole organism is referred to as "reproductive cloning." Long before attempts were made to clone an entire organism, researchers learned how to reproduce desired regions or fragments of the genome, a process that is referred to as molecular cloning.

Cloning small genome fragments allows researchers to manipulate and study specific genes (and their protein products), or noncoding regions in isolation. A plasmid, or vector, is a small circular DNA molecule that replicates independently of the chromosomal DNA. In cloning, scientists can use the plasmid molecules to provide a "folder" in which to insert a desired DNA fragment. Plasmids are usually introduced into a bacterial host for proliferation. In the bacterial context, scientists call the DNA fragment from the human genome (or the genome of another studied organism) **foreign DNA**, or a transgene, to differentiate it from the bacterium's DNA, or the **host DNA**.

Plasmids occur naturally in bacterial populations (such as Escherichia coli) and have genes that can contribute favorable traits to the organism, such as **antibiotic resistance** (the ability to be unaffected by antibiotics). Scientists have repurposed and engineered plasmids as vectors for molecular cloning and the large-scale production of important reagents, such as insulin and human growth hormone. An important feature of plasmid vectors is the ease with which scientists can introduce a foreign DNA fragment via the **multiple cloning site (MCS)**. The MCS is a short DNA sequence containing multiple sites that different commonly available restriction endonucleases can cut. **Restriction endonucleases** recognize specific DNA sequences and cut them in a predictable manner. They are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction

endonucleases make staggered cuts in the two DNA strands, such that the cut ends have a 2- or 4-base singlestranded overhang. Because these overhangs are capable of annealing with complementary overhangs, we call them "sticky ends." Adding the enzyme DNA ligase permanently joins the DNA fragments via phosphodiester bonds. In this way, scientists can splice any DNA fragment generated by restriction endonuclease cleavage between the plasmid DNA's two ends that has been cut with the same restriction endonuclease (Figure 53).

Recombinant DNA Molecules

Plasmids with foreign DNA inserted into them are called **recombinant DNA** molecules because they are created artificially and do not occur in nature. They are also called chimeric molecules because the origin of different molecule parts of the molecules can be traced back to different species of biological organisms or even to chemical synthesis. We call proteins that are expressed from recombinant DNA molecules **recombinant proteins**. Not all recombinant plasmids are capable of expressing genes. The recombinant DNA may need to move into a different vector (or host) that is better designed for gene expression. Scientists may also engineer plasmids to express proteins only when certain environmental factors stimulate them, so they can control therecombinant proteins' expression.

Cellular Cloning

Unicellular organisms, such as bacteria and yeast, naturally produce clones of themselves when they replicate asexually by binary fission; this is known as **cellular cloning**. The nuclear DNA duplicates by the process of mitosis, which creates an exact replica of the genetic material.

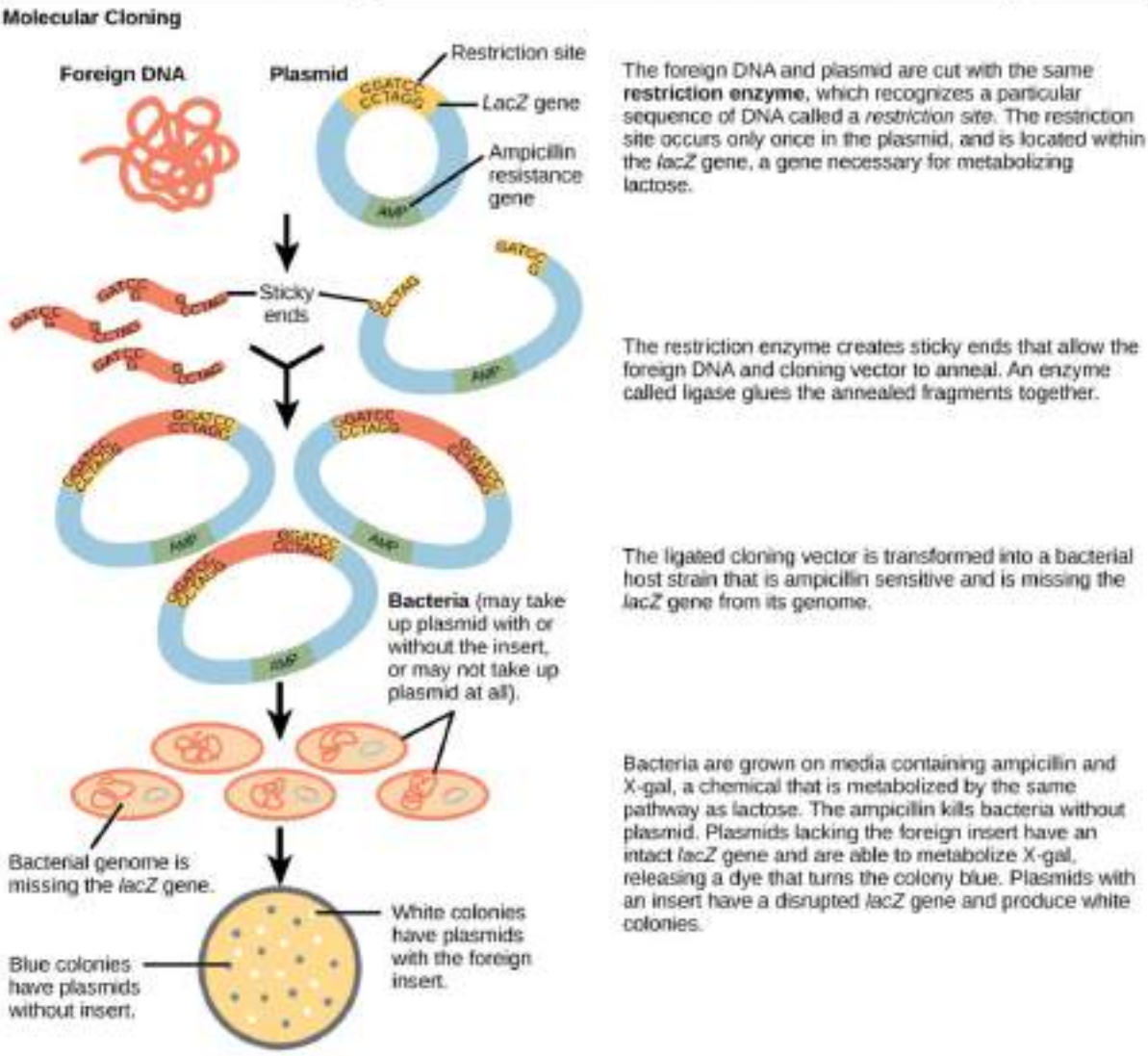


Figure 53: This diagram shows the steps involved in molecular cloning

Reproductive Cloning

Reproductive cloning is a method scientists use to clone or identically copy an entire multicellular organism. Most multicellular organisms undergo reproduction by sexual means, which involves genetic hybridization of two individuals (parents), making it impossible to generate an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to artificially induce mammal asexual reproduction in the laboratory.

Parthenogenesis, or "virgin birth," occurs when an embryo grows and develops without egg fertilization. This is a form of asexual reproduction. An example of parthenogenesis occurs in species in which the female lays an egg and if the egg is fertilized, it is a diploid egg and the individual develops into a female. If the egg is not fertilized, it remains a haploid egg and develops into a male. The unfertilized egg is a parthenogenic, or virgin egg. Some insects and reptiles lay parthenogenic eggs that can develop into adults.

Sexual reproduction requires two cells. When the haploid egg and sperm cells fuse, a diploid zygote results. The zygote nucleus contains the genetic information to produce a new individual. However, early embryonic development requires the cytoplasmic material contained in the egg cell. This idea forms the basis for reproductive cloning. Therefore, if we replace the egg cell's haploid nucleus with a diploid nucleus from the cell of any individual of the same species (a donor), it will become a zygote that is genetically identical to the donor. Somatic cell nuclear transfer is the technique of transferring a diploid nucleus into an enucleated egg. Scientists can use it for either therapeutic cloning or reproductive cloning.

The first cloned animal was Dolly, a sheep born in 1996. The reproductive cloning success rate at the time was very low. Dolly lived for seven years and died of respiratory complications (Figure 54). There is speculation that because the cell DNA belongs to an older individual, DNA's age may affect a cloned individual's life expectancy. Since Dolly, scientists have cloned successfully several animals such as horses, bulls, and goats, although these animals often exhibit facial, limb, and cardiac abnormalities. There have been attempts at producing cloned human embryos as sources of embryonic stem cells for therapeutic purposes. Therapeutic cloning produces stem cells in the attempt to remedy detrimental diseases or defects (unlike reproductive cloning, which aims to reproduce an organism). Still, some

have met therapeutic cloning efforts with resistance because of bioethical considerations.

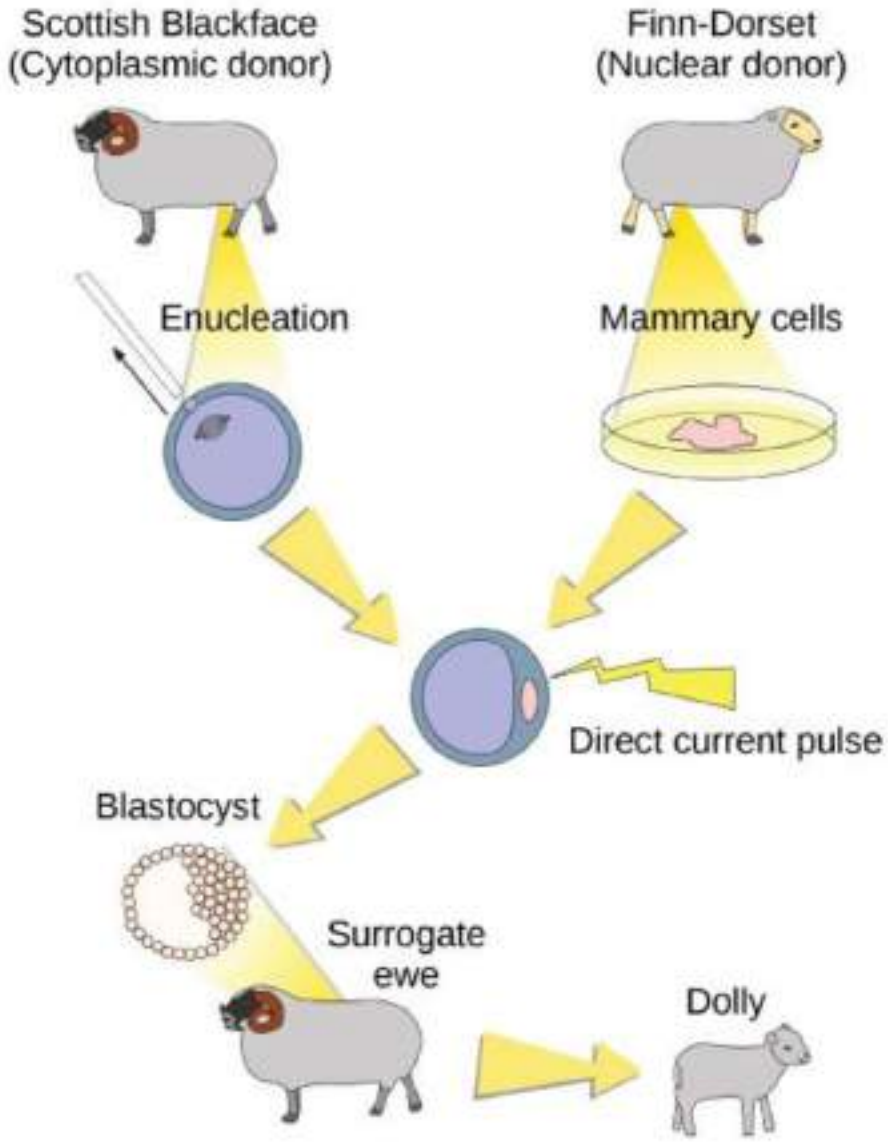


Figure 54: Dolly the sheep was the first mammal to be cloned. To create Dolly, they removed the nucleus from a donor egg cell. They then introduced the nucleus from a second sheep into the cell, which divided to the blastocyst stage before they implanted it in a surrogate mother. (credit: modification of work by "Squidonius"/Wikimedia Commons)

Genetic Engineering

Genetic engineering is the alteration of an organism's genotype using recombinant DNA technology to modify an organism's DNA to achieve desirable traits. The addition of foreign DNA in the form of recombinant DNA vectors generated by molecular cloning is the most common method of genetic engineering. The organism that receives the recombinant DNA is a **genetically modified organism** (GMO). If the foreign DNA comes from a different species, the host organism is **transgenic**. Scientists have genetically modified bacteria, plants, and animals since the early 1970s for academic, medical, agricultural, and industrial purposes. In the US, GMOs such as Roundup-ready soybeans and borer-resistant corn are part of many common processed foods.

Gene Targeting

Although classical methods of studying gene function began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: "What does this gene or DNA element do?" This technique, reverse genetics, has resulted in reversing the classic genetic methodology. This method would be similar to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the wing's function is flight. The classical genetic method would compare insects that cannot fly with insects that can fly, and observe that the non-flying insects have lost wings. Similarly, mutating or deleting genes provides researchers with clues about gene function. We collectively call the methods they use to disable gene function gene targeting. **Gene targeting** is the use of recombinant DNA vectors to alter a particular gene's expression, either by introducing mutations in a gene, or by eliminating a certain gene's expression by deleting a part or all of the gene sequence from the organism's genome.

Biotechnology in Medicine and Agriculture

It is easy to see how biotechnology can be used for medicinal purposes. Knowledge of the genetic makeup of our species, the genetic basis of heritable diseases, and the invention of technology to manipulate and fix mutant genes provides methods to treat the disease. Biotechnology in agriculture can enhance resistance to disease, pest, and environmental stress, and improve both crop yield and quality.

Genetic Diagnosis and Gene Therapy

Scientists call the process of testing for suspected genetic defects before administering treatment **genetic diagnosis** by **genetic testing**. Depending on the inheritance patterns of a disease-causing gene, family members are advised to undergo genetic testing. For example, doctors usually advise women diagnosed with breast cancer to have a biopsy so that the medical team can determine the genetic basis of cancer development. Doctors base treatment plans on genetic test findings that determine the type of cancer. If inherited gene mutations cause the cancer, doctors also advise other female relatives to undergo genetic testing and periodic screening for breast cancer. Doctors also offer genetic testing for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

Gene therapy is a genetic engineering technique used to cure disease. In its simplest form, it involves the introduction of a good gene at a random location in the genome to aid the cure of a disease that is caused by a mutated gene. The good gene is usually introduced into diseased cells as part of a vector transmitted by a virus that can infect the host cell and deliver the foreign DNA (Figure 55). More advanced forms of gene therapy try to correct the mutation at the original site in the genome, such as is the case with treatment of severe combined immunodeficiency (SCID).

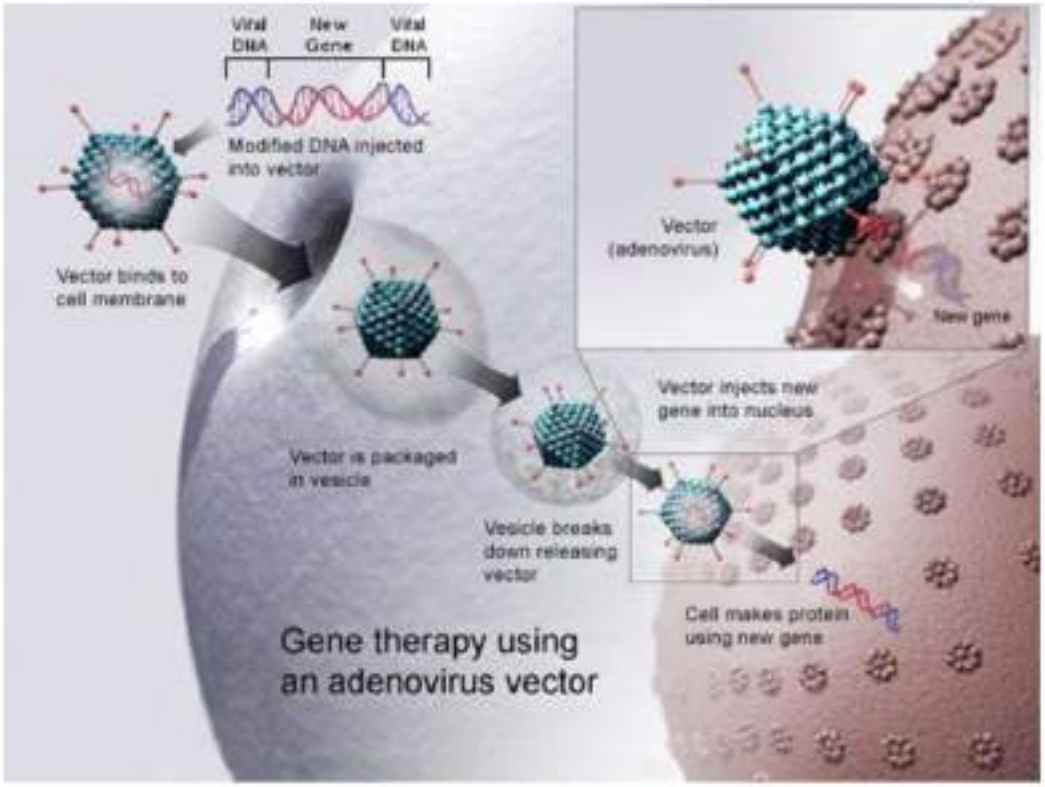


Figure 55: Gene therapy using an adenovirus vector can be used to cure certain genetic diseases in which a person has a defective gene. (credit: NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms to mount the initial immune response. Modern techniques use the genes of microorganisms cloned into vectors to mass produce the desired antigen. Doctors then introduce the antigen into the body to stimulate the primary immune response and trigger immune memory. The medical field has used genes cloned from the influenza virus to combat the constantly changing strains of this virus.

Antibiotics are a biotechnological product. Microorganisms, such as fungi, naturally produce them to attain an advantage over bacterial populations. Cultivating and manipulating fungal cells produces antibodies.

Scientists used recombinant DNA technology to produce large-scale quantities of

human insulin in E. coli as early as 1978. Previously, it was only possible to treat diabetes with pig insulin, which caused allergic reactions in humans because of differences in the gene product. In addition, doctors use human growth hormone (HGH) to treat growth disorders in children. Researchers cloned the HGH gene from a cDNA library and inserted it into E. coli cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins in medicine are successfully produced in bacteria, some proteins require a eukaryotic animal host for proper processing. For this reason, the desired genes are cloned and expressed in animals, such as sheep, goats, chickens, and mice. We call animals that have been modified to express recombinant DNA transgenic animals. Several human proteins are expressed in transgenic sheep and goat milk, and some are expressed in chicken eggs. Scientists have used mice extensively for expressing and studying recombinant gene and mutation effects.

Transgenic Plants

Manipulating the DNA of plants (i.e., creating GMOs) has helped to create desirable traits, such as disease resistance, herbicide and pesticide resistance, better nutritional value, and better shelf-life (Figure 56). Plants are the most important source of food for the human population. Farmers developed ways to select for plant varieties with desirable traits long before moderndday biotechnology practices were established.

Figure 56: Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

We call plants that have received recombinant DNA from other species transgenic plants. Because they are not natural, government agencies closely monitor transgenic plants and other GMOs to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because foreign genes can spread to other species in the environment, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants that scientists genetically engineered.

Transformation of Plants Using Agrobacterium tumefaciens

Gene transfer occurs naturally between species in microbial populations. Many viruses that cause human diseases, such as cancer, act by incorporating their DNA into the human genome. In plants, tumors caused by the bacterium Agrobacterium tumefaciens occur by DNA transfer from the bacterium to the plant. Although the tumors do not kill the plants, they stunt the plants and they become more susceptible to harsh environmental conditions. A. tumefaciens affects many plants such as walnuts, grapes, nut trees, and beets. Artificially introducing DNA into plant cells is more challenging than in animal cells because of the thick plant cell wall.

Researchers used the natural transfer of DNA from Agrobacterium to a plant host to introduce DNA fragments of their choice into plant hosts. In nature, the disease-causing A. tumefaciens have a set of plasmids, **Ti plasmids** (tumor-inducing plasmids), that contain genes to produce tumors in plants. DNA from the Ti plasmid integrates into the infected plant cell's genome. Researchers manipulate the Ti plasmids to remove the tumor-causing genes and insert the desired DNA fragment for transfer into the plant genome. The Ti plasmids carry antibiotic resistance genes to aid selection and researchers can propagate them in E. coli cells as well.

The Organic Insecticide Bacillus thuringiensis

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals during sporulation that are toxic to many insect species that affect plants. Insects need to ingest Bt toxin in order to activate the toxin. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin activates in the insects' intestines, they die within a couple of days. Modern biotechnology has allowed plants to encode their own crystal Bt toxin that acts against insects. Scientists have cloned the crystal toxin genes from Bt and introduced them into plants. Bt toxin is

safe for the environment, nontoxic to humans and other mammals, and organic farmers have approved it as a natural insecticide.

Flavr Savr Tomato

The first GM crop on the market was the Flavr Savr Tomato in 1994. Scientists used antisense RNA technology to slow the softening and rotting process caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the tomato's flavor. The Flavr Savr tomato did not successfully stay in the market because of problems maintaining and shipping the crop.

Mapping Genomes

Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species. **Genome mapping** is the process of finding the locations of genes on each chromosome. The maps that genome mapping create are comparable to the maps that we use to navigate streets. A **genetic map** is an illustration that lists genes and their location on a chromosome. Genetic maps provide the big picture (similar to an interstate highway map) and use genetic markers (similar to landmarks). A **genetic marker** is a gene or sequence on a chromosome that co-segregates (shows genetic linkage) with a specific trait. Early geneticists called this linkage analysis. Physical maps present the intimate details of smaller chromosome regions (similar to a detailed road map). A **physical map** is a representation of the physical distance, in nucleotides, between genes or genetic markers. Both genetic linkage maps and physical maps are required to build a genome's complete picture. Having a complete genome map of the genome makes it easier for researchers to study individual genes. Human genome maps help researchers in their efforts to identify human

diseasecausing genes related to illnesses like cancer, heart disease, and cystic fibrosis. We can use genome mapping in a variety of other applications, such as using live microbes to clean up pollutants or even prevent pollution. Research involving plant genome mapping may lead to producing higher crop yields or developing plants that better adapt to climate change.

Genetic Maps

The study of genetic maps begins with **linkage analysis**, a procedure that analyzes the recombination frequency between genes to determine if they are linked or show independent assortment. Scientists used the term linkage before the discovery of DNA. Early geneticists relied on observing phenotypic changes to understand an organism's genotype. Shortly after Gregor Mendel (the father of modern genetics) proposed that traits were determined by what we now call genes, other researchers observed that different traits were often inherited together, and thereby deduced that the genes were physically linked by their location on the same chromosome. Gene mapping relative to each other based on linkage analysis led to developing the first genetic maps.

Observations that certain traits were always linked and certain others were not linked came from studying the offspring of crosses between parents with different traits. For example, in garden pea experiments, researchers discovered, that the flower's color and plant pollen's shape were linked traits, and therefore the genes encoding these traits were in close proximity on the same chromosome. We call exchanging DNA between homologous chromosome pairs **genetic recombination**, which occurs by crossing over DNA between homologous DNA strands, such as nonsister chromatids. Linkage analysis involves studying the recombination frequency between any two genes. The greater the distance between two genes, the higher the

chance that a recombination event will occur between them, and the higher the recombination frequency between them. Figure 57 shows two possibilities for recombination between two nonsister chromatids during meiosis. If the recombination frequency between two genes is less than 50 percent, they are linked.

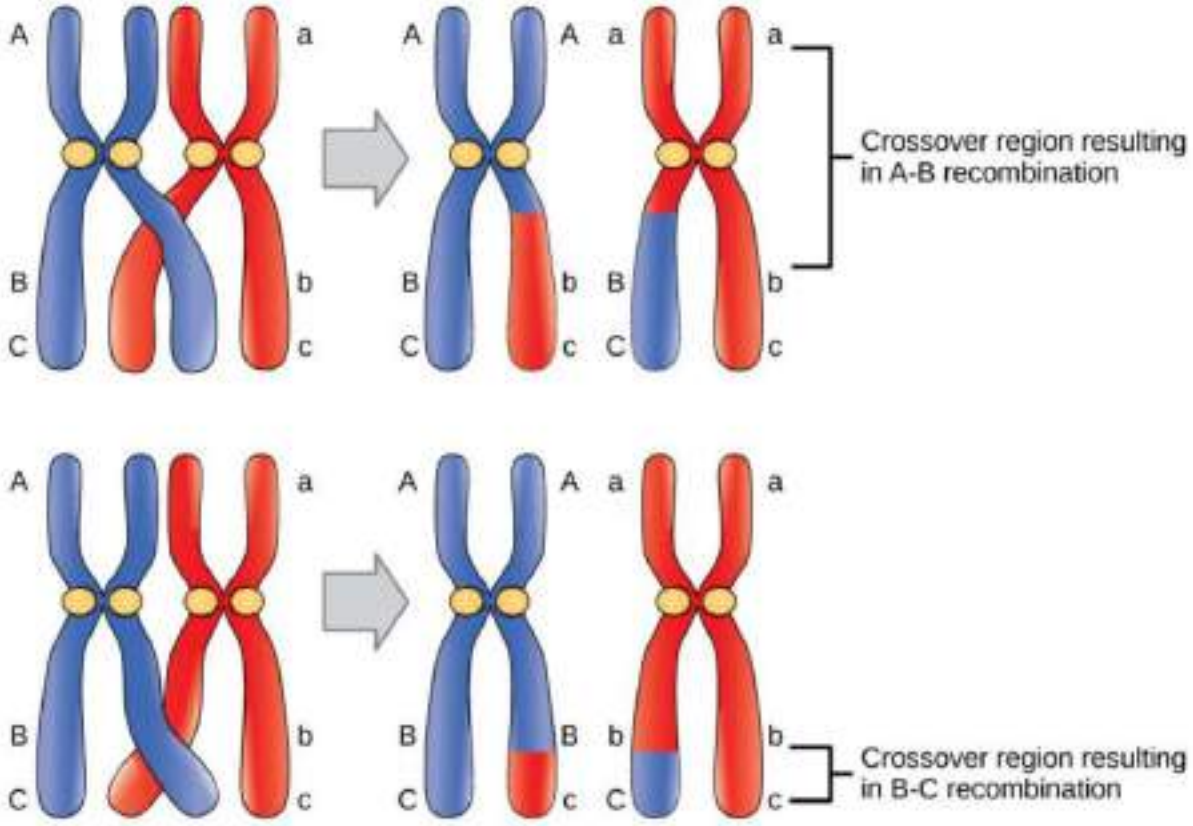


Figure 57: Crossover may occur at different locations on the chromosome. Recombination between genes A and B is more frequent than recombination between genes B and C because genes A and B are farther apart. Therefore, a crossover is more likely to occur between them.

The generation of genetic maps requires markers, just as a road map requires landmarks (such as rivers and mountains). Scientists based early genetic maps on using known genes as markers. Scientists now use more sophisticated markers, including those based on non-coding DNA, to compare individuals' genomes in a population. Although individuals of a given species are genetically similar, they are

not identical. Every individual has a unique set of traits. These minor differences in the genome between individuals in a population are useful for genetic mapping purposes. In general, a good genetic marker is a region on the chromosome that shows variability or polymorphism (multiple forms) in the population.

Some genetic markers that scientists use in generating genetic maps are **restriction fragment length polymorphisms** (RFLP), variable number of tandem repeats (VNTRs), **microsatellite polymorphisms**, and the **single nucleotide polymorphisms** (SNPs). We can detect RFLPs (sometimes pronounced "rif-lips") when the DNA of an individual is cut with a restriction endonuclease that recognizes specific sequences in the DNA to generate a series of DNA fragments, which we can then analyze using gel electrophoresis. Every individual's DNA will give rise to a unique pattern of bands when cut with a particular set of restriction endonucleases. Scientists sometimes refer to this as an individual's DNA "fingerprint." Certain chromosome regions that are subject to polymorphism will lead to generating the unique banding pattern. VNTRs are repeated sets of nucleotides present in DNA's non-coding regions. Non-coding, or "junk," DNA has no known biological function; however, research shows that much of this DNA is actually transcribed. While its function is uncertain, it is certainly active, and it may be involved in regulating coding genes. The number of repeats may vary in a population's individual organisms. Microsatellite polymorphisms are similar to VNTRs, but the repeat unit is very small. SNPs are variations in a single nucleotide.

Because genetic maps rely completely on the natural process of recombination, natural increases or decreases in the recombination level given genome area affects mapping. Some parts of the genome are recombination hotspots; whereas, others do not show a propensity for recombination. For this reason, it is important to look at mapping information developed by multiple methods.

Physical Maps

A physical map provides detail of the actual physical distance between genetic markers, as well as the number of nucleotides. There are three methods scientists use to create a physical map: cytogenetic mapping, radiation hybrid mapping, and sequence mapping. **Cytogenetic mapping** uses information from microscopic analysis of stained chromosome sections (Figure 58). It is possible to determine the approximate distance between genetic markers using cytogenetic mapping, but not the exact distance (number of base pairs). **Radiation hybrid mapping** uses radiation, such as x-rays, to break the DNA into fragments. We can adjust the radiation amount to create smaller or larger fragments. This technique overcomes the limitation of genetic mapping, and we can adjust the radiation so that increased or decreased recombination frequency does not affect it. **Sequence mapping** resulted from DNA sequencing technology that allowed for creating detailed physical maps with distances measured in terms of the number of base pairs. Creating **genomic libraries** and **complementary DNA (cDNA) libraries** (collections of cloned sequences or all DNA from a genome) has sped the physical mapping process. A genetic site that scientists use to generate a physical map with sequencing technology (a sequence-tagged site, or STS) is a unique sequence in the genome with a known exact chromosomal location. An **expressed sequence tag (EST)** and a single sequence length polymorphism (SSLP) are common STSs. An EST is a short STS that we can identify with cDNA libraries, while we obtain SSLPs from known genetic markers, which provide a link between genetic and physical maps.

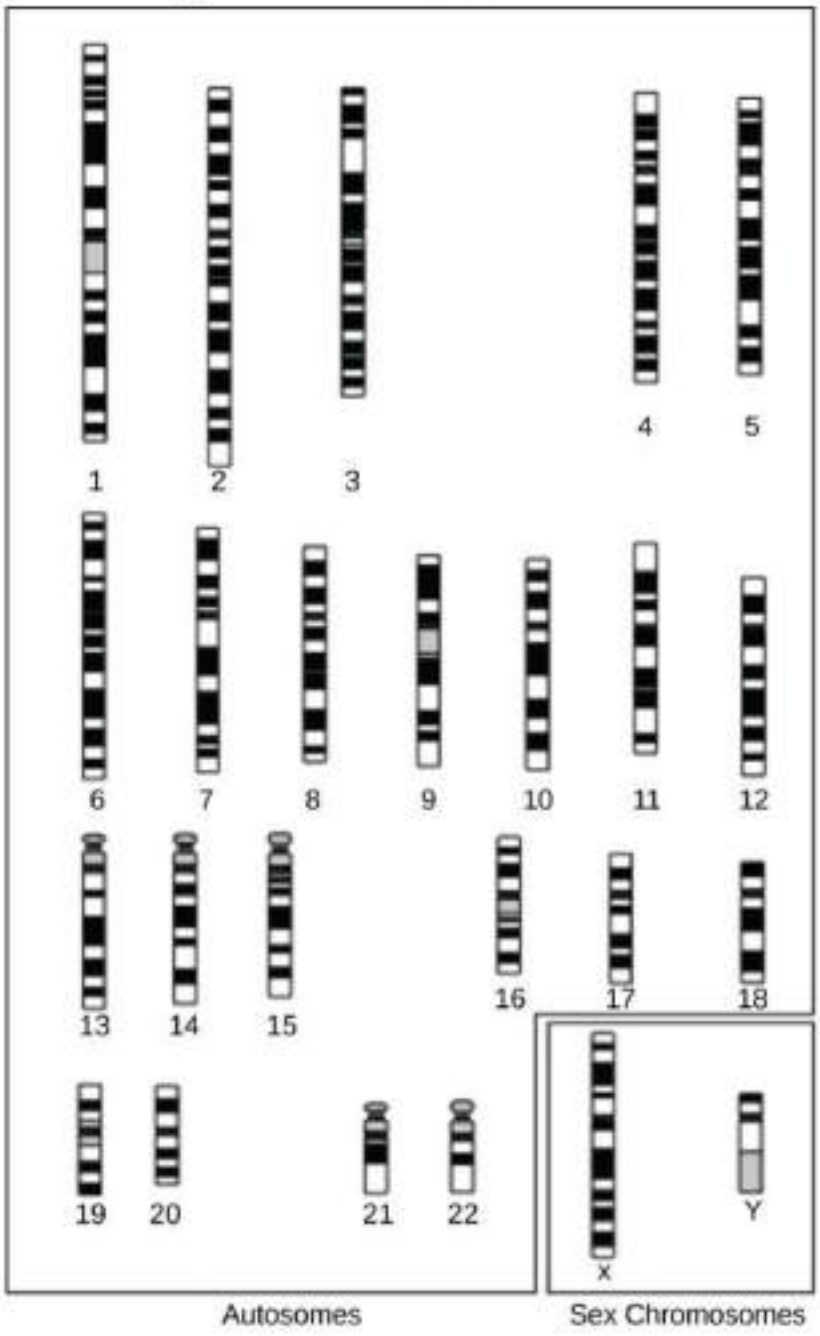


Figure 58: A cytogenetic map shows the appearance of a chromosome after scientists stain and exam it under a microscope. (credit: National Human Genome Research Institute)

Genetic and Physical Maps Integration

Genetic maps provide the outline and physical maps provide the details. It is easy to understand why both genome mapping technique types are important to show the big picture. Scientists use information from each technique in combination to study the genome. Scientists are using genomic mapping with different model organisms for research. Genome mapping is still an ongoing process, and as researchers develop more advanced techniques, they expect more breakthroughs. Genome mapping is similar to completing a complicated puzzle using every piece of available data. Mapping information generated in laboratories all over the world goes into central databases, such as GenBank at the National Center for Biotechnology Information (NCBI).

Researchers are making efforts for the information to be more easily accessible to other researchers and the general public. Just as we use global positioning systems instead of paper maps to navigate through roadways, NCBI has created a genome viewer tool to simplify the data-mining process

Whole-Genome Sequencing

Although there have been significant advances in the medical sciences in recent years, doctors are still confounded by some diseases, and they are using whole-genome sequencing to discover the root of the problem. **Whole-genome sequencing** is a process that determines an entire genome's DNA sequence. Whole-genome sequencing is a brute-force approach to problem solving when there is a genetic basis at the core of a disease. Several laboratories now provide services to sequence, analyze, and interpret entire genomes.

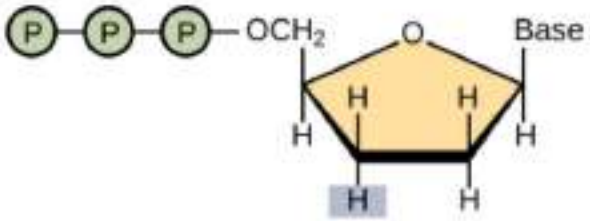
For example, whole-exome sequencing is a lower-cost alternative to whole genome sequencing. In exome sequencing, the doctor sequences only the DNA's coding,

exon-producing regions. In 2010, doctors used whole-exome sequencing to save a young boy whose intestines had multiple mysterious abscesses. The child had several colon operations with no relief. Finally, they performed whole-exome sequencing, which revealed a defect in a pathway that controls apoptosis (programmed cell death). The doctors used a bone-marrow transplant to overcome this genetic disorder, leading to a cure for the boy. He was the first person to receive successful treatment based on a whole-exome sequencing diagnosis. Today, human genome sequencing is more readily available and results are available within two days for about $1000.

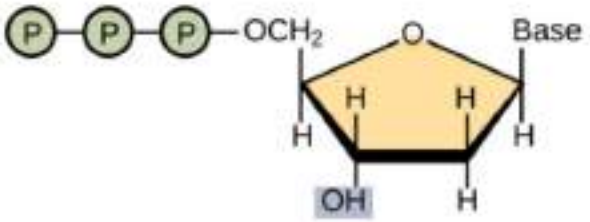
Strategies Used in Sequencing Projects

The basic sequencing technique used in all modern day sequencing projects is the chain termination method (also known as the dideoxy method), which Fred Sanger developed in the 1970s. The chain termination method involves DNA replication of a singlestranded template by using a primer and a regular **deoxynucleotide** (dNTP), which is a monomer, or a single DNA unit. The primer and dNTP mix with a small proportion of fluorescently labeled **dideoxynucleotides** (ddNTPs). The ddNTPs are monomers that are missing a hydroxyl group (–OH) at the site at which another nucleotide usually attaches to form a chain (Figure 59). Scientists label each ddNTP with a different color of fluorophore. Every time a ddNTP incorporates in the growing complementary strand, it terminates the DNA replication process, which results in multiple short strands of replicated DNAthat each terminate at a different point during replication. When gel electrophoresis processes the reaction mixture after separating into single strands, the multiple newly replicated DNA strands form a ladder because of the differing sizes. Because the ddNTPs are fluorescently labeled, each band on the gel reflects the DNA strand's size and the ddNTP that

terminated the reaction. The different colors of the fluorophore-labeled ddNTPs help identify the ddNTP incorporated at that position. Reading the gel on the basis of each band's color on the ladder produces the template strand's sequence (Figure 60).



Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

Figure 59: A dideoxynucleotide is similar in structure to a deoxynucleotide, but is missing the 3' hydroxyl group (indicated by the box). When a dideoxynucleotide is incorporated into a DNA strand, DNA synthesis stops

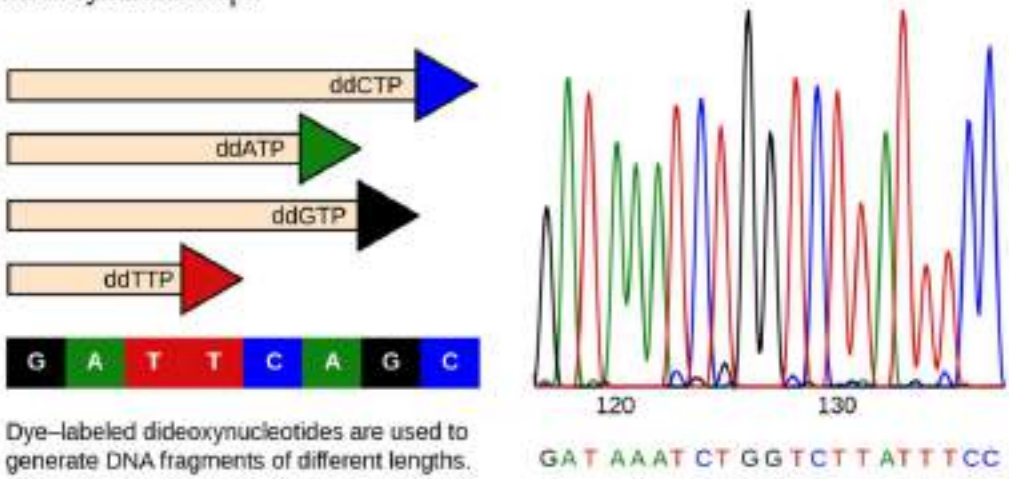


Figure 60: This figure illustrates Frederick Sanger's dideoxy chain termination method. Using dideoxynucleotides, the DNA fragment can terminate at different points. The DNA separates on the basis of size, and we can read these bands based on the fragments' size.

Early Strategies: Shotgun Sequencing and Pair-Wise End Sequencing

In **shotgun sequencing** method, several DNA fragment copies cut randomly into many smaller pieces (somewhat like what happens to a round shot cartridge when fired from a shotgun). All of the segments sequence using the chain-sequencing method. Then, with sequence computer assistance, scientists can analyze the fragments to see where their sequences overlap. By matching overlapping sequences at each fragment's end, scientists can reform the entire DNA sequence. A larger sequence that is assembled from overlapping shorter sequences is called a **contig**. As an analogy, consider that someone has four copies of a landscape photograph that you have never seen before and know nothing about how it should appear. The person then rips up each photograph with their hands, so that different size pieces are present from each copy. The person then mixes all of the pieces together and asks you to reconstruct the photograph. In one of the smaller pieces you see a mountain. In a larger piece, you see that the same mountain is behind a lake. A third fragment shows only the lake, but it reveals that there is a cabin on the shore of the lake. Therefore, from looking at the overlapping information in these three fragments, you know that the picture contains a mountain behind a lake that has a cabin on its shore. This is the principle behind reconstructing entire DNA sequences using shotgun sequencing. Originally, shotgun sequencing only analyzed one end of each fragment for overlaps. This was sufficient for sequencing small genomes. However, the desire to sequence larger genomes, such as that of a human, led to developing double-barrel shotgun sequencing, or **pairwise-end sequencing**. In pairwise-end sequencing, scientists analyze each fragment's end for overlap. Pairwise-end sequencing is, therefore, more cumbersome than shotgun sequencing, but it is easier to reconstruct the sequence because there is more available information.

Next-generation Sequencing

Since 2005, automated sequencing techniques used by laboratories are under the umbrella of **next-generation sequencing**, which is a group of automated techniques used for rapid DNA sequencing. These automated low-cost sequencers can generate sequences of hundreds of thousands or millions of short fragments (25 to 500 base pairs) in the span of one day. These sequencers use sophisticated software to get through the cumbersome process of putting all the fragments in order.

Comparing Sequences

A sequence alignment is an arrangement of proteins, DNA, or RNA. Scientists use it to identify similar regions between cell types or species, which may indicate function or structure conservation. We can use sequence alignments to construct phylogenetic trees. The following website uses a software program called BLAST (basic local alignment search tool) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) .

Under "Basic Blast," click "Nucleotide Blast." Input the following sequence into the large "query sequence" box: ATTGCTTCGATTGCA. Below the box, locate the "Species" field and type "human" or "Homo sapiens". Then click "BLAST" to compare the inputted sequence against the human genome's known sequences. The result is that this sequence occurs in over a hundred places in the human genome. Scroll down below the graphic with the horizontal bars and you will see a short description of each of the matching hits. Pick one of the hits near the top of the list and click on "Graphics". This will bring you to a page that shows the sequence's location within the entire human genome. You can move the slider that looks like a green flag back and forth to view the sequences immediately around the selected gene. You can then return to your selected sequence by clicking the "ATG" button.

Use of Whole-Genome Sequences of Model Organisms

British biochemist and Nobel Prize winner Fred Sanger used a bacterial virus, the bacteriophage fx174 (5368 base pairs), to completely sequence the first genome. Other scientists later sequenced several other organelle and viral genomes. American biotechnologist, biochemist, geneticist, and businessman Craig Venter sequenced the bacterium Haemophilus influenzae in the 1980s. Approximately 74 different laboratories collaborated on sequencing the genome of the yeast Saccharomyces cerevisiae, which began in 1989 and was completed in 1996, because it was 60 times bigger than any other genome sequencing. By 1997, the genome sequences of two important model organisms were available: the bacterium Escherichia coli K12 and the yeast Saccharomyces cerevisiae. We now know the genomes of other model organisms, such as the mouse Mus musculus, the fruit fly Drosophila melanogaster, the nematode Caenorhabditis. elegans, and humans Homo sapiens. Researchers perform extensive basic research in model organisms because they can apply the information to genetically similar organisms. A **model organism** is a species that researchers use as a model to understand the biological processes in other species that the model organism represents. Having entire genomes sequenced helps with the research efforts in these model organisms. The process of attaching biological information to gene sequences is **genome annotation**. Annotating gene sequences helps with basic experiments in molecular biology, such as designing PCR primers and RNA targets.

Genome Sequence Uses

DNA microarrays are methods that scientists use to detect gene expression by analyzing different DNA fragments that are fixed to a glass slide or a silicon chip to identify active genes and sequences. We can discover almost one million genotypic

abnormalities using microarrays; whereas, whole-genome sequencing can provide information about all six billion base pairs in the human genome. Although studying genome sequencing medical applications is interesting, this discipline dwells on abnormal gene function. Knowing about the entire genome will allow researchers to discover future onset diseases and other genetic disorders early. This will allow for more informed decisions about lifestyle, medication, and having children. Genomics is still in its infancy, although someday it may become routine to use whole-genome sequencing to screen every newborn to detect genetic abnormalities.

In addition to disease and medicine, genomics can contribute to developing novel enzymes that convert biomass to biofuel, which results in higher crop and fuel production, and lower consumer cost. This knowledge should allow better methods of control over the microbes that industry uses to produce biofuels. Genomics could also improve monitoring methods that measure the impact of pollutants on ecosystems and help clean up environmental contaminants. Genomics has aided in developing agrochemicals and pharmaceuticals that could benefit medical science and agriculture.

It sounds great to have all the knowledge we can get from whole-genome sequencing; however, humans have a responsibility to use this knowledge wisely. Otherwise, it could be easy to misuse the power of such knowledge, leading to discrimination based on a person's genetics, human genetic engineering, and other ethical concerns. This information could also lead to legal issues regarding health and privacy.

Applying Genomics

Introducing DNA sequencing and whole genome sequencing projects, particularly the Human Genome project, has expanded the applicability of DNA sequence information. Many fields, such as metagenomics, pharmacogenomics, and mitochondrial genomics are using genomics. Understanding and finding cures for diseases is the most common application of genomics.

Predicting Disease Risk at the Individual Level

Predicting disease risk involves screening currently healthy individuals by genome analysis at the individual level. Health care professionals can recommend intervention with lifestyle changes and drugs before disease onset. However, this approach is most applicable when the problem resides within a single gene defect. Such defects only account for approximately 5 percent of diseases in developed countries. Most of the common diseases, such as heart disease, are multi-factored or **polygenic**, which is a phenotypic characteristic that involves two or more genes, and also involve environmental factors such as diet. In April 2010, scientists at Stanford University published the genome analysis of a healthy individual (Stephen Quake, a scientist at Stanford University, who had his genome sequenced. The analysis predicted his propensity to acquire various diseases. The medical team performed a risk assessment to analyze Quake's percentage of risk for 55 different medical conditions. The team found a rare genetic mutation, which showed him to be at risk for sudden heart attack. The results also predicted that Quake had a 23 percent risk of developing prostate cancer and a 1.4 percent risk of developing Alzheimer's. The scientists used databases and several publications to analyze the genomic data. Even though genomic sequencing is becoming more affordable and analytical tools are becoming more reliable, researchers still must address ethical issues surrounding genomic analysis at a population level.

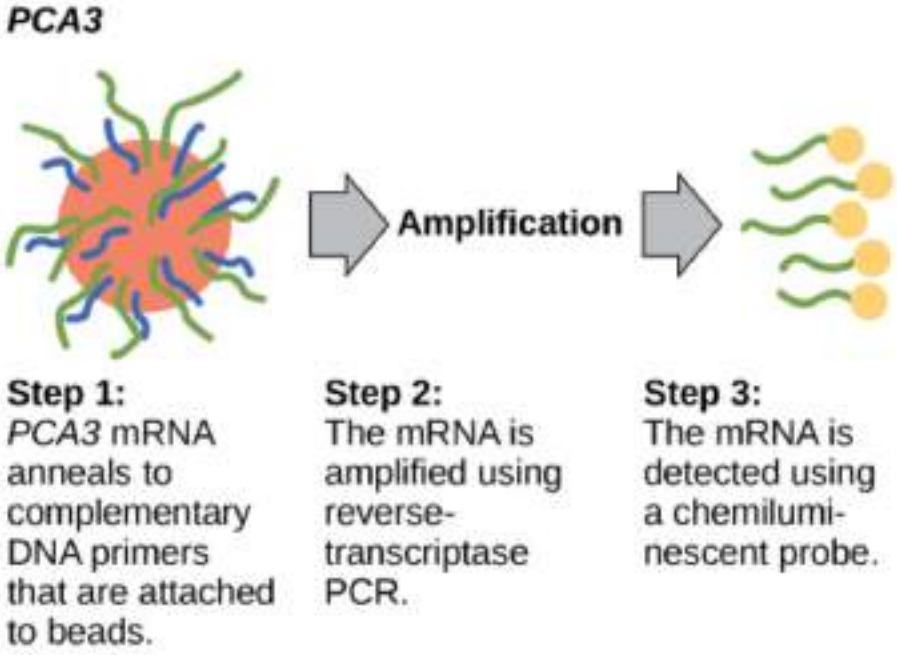


Figure 61: PCA3 is a gene that is expressed in prostate epithelial cells and overexpressed in cancerous cells. A high PCA3 concentration in urine is indicative of prostate cancer. The PCA3 test is a better indicator of cancer than the more well known PSA test, which measures the level of PSA (prostate-specific antigen) in the blood.

In 2011, the United States Preventative Services Task Force recommended against using the PSA test to screen healthy men for prostate cancer. Their recommendation is based on evidence that screening does not reduce the risk of death from prostate cancer. Prostate cancer often develops very slowly and does not cause problems, while the cancer treatment can have severe side effects. The PCA3 test is more accurate, but screening may still result in men who would not have been harmed by the cancer itself suffering side effects from treatment. What do you think? Should all healthy men receive prostate cancer screenings using the PCA3 or PSA test? Should people in general receive screenings to find out if they have a genetic risk for cancer or other diseases?

Pharmacogenomics and Toxicogenomics

Pharmacogenomics, or toxicogenomics, involves evaluating drug effectiveness and safety on the basis of information from an individual's genomic sequence. We can study genomic responses to drugs using experimental animals (such as laboratory rats or mice) or live cells in the laboratory before embarking on studies with humans. Studying changes in gene expression could provide information about the transcription profile in the drug's presence, which we can use as an early indicator of the potential for toxic effects. For example, genes involved in cellular growth and controlled cell death, when disturbed, could lead to cancerous cell growth. Genome-wide studies can also help to find new genes involved in drug toxicity. Medical professionals can use personal genome sequence information to prescribe medications that will be most effective and least toxic on the basis of the individual patient's genotype. The gene signatures may not be completely accurate, but medical professionals can test them further before pathologic symptoms arise.

Microbial Genomics: Metagenomics

Traditionally, scholars have taught microbiology with the view that it is best to study microorganisms under **pure culture** conditions. This involves isolating a single cell type and culturing it in the laboratory. Because microorganisms can go through several generations in a matter of hours, their gene expression profiles adapt to the new laboratory environment very quickly. In addition, the vast majority of bacterial species resist culturing in isolation. Most microorganisms do not live as isolated entities, but in microbial communities or biofilms. For all of these reasons, pure culture is not always the best way to study microorganisms. **Metagenomics** is the study of the collective genomes of multiple species that grow and interact in an environmental niche. Metagenomics can be used to identify new species more rapidly and to analyze the effect of pollutants on the environment (Figure 62).

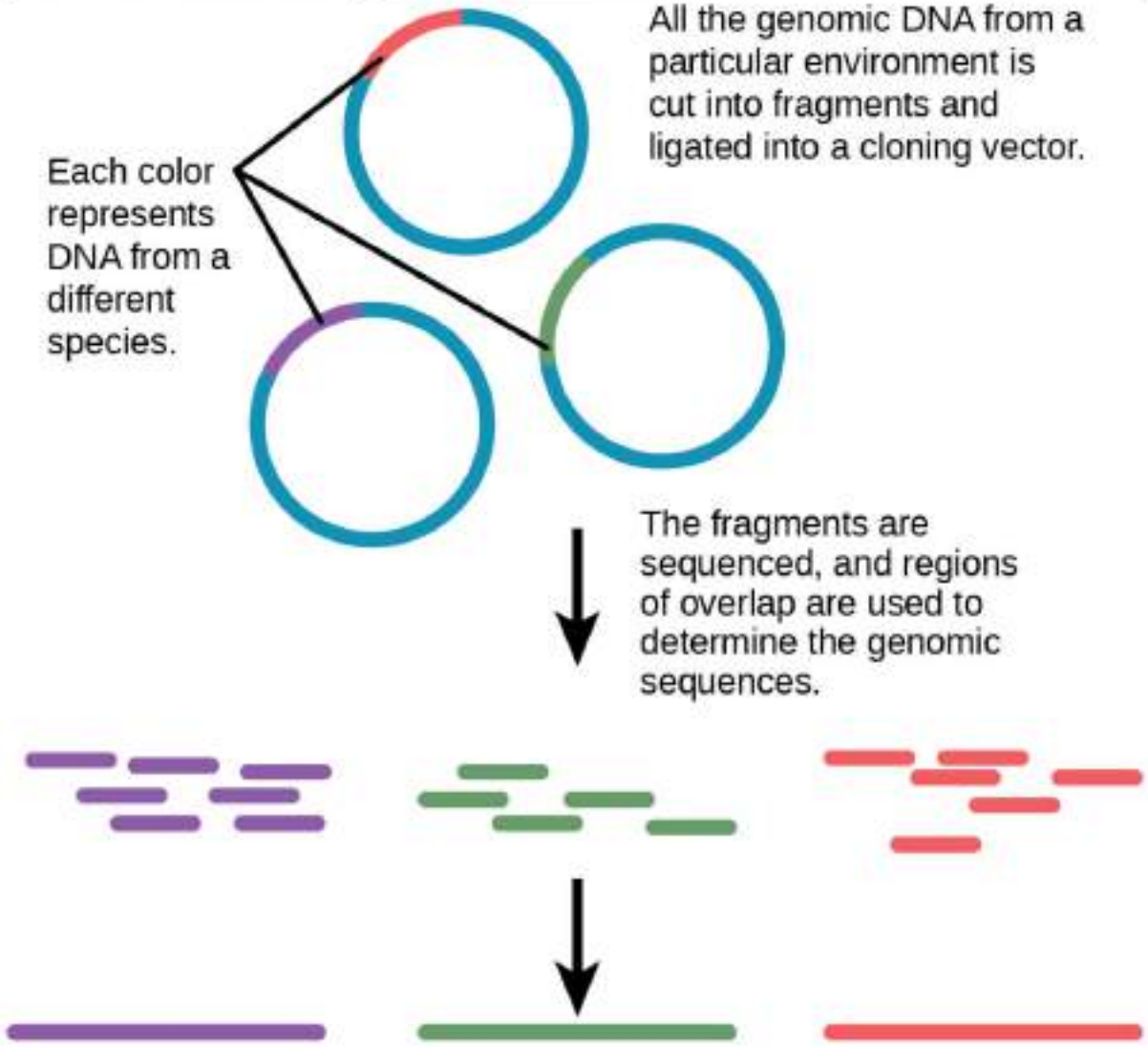


Figure 62: Metagenomics involves isolating DNA from multiple species within an environmental niche

Microbial Genomics: Creation of New Biofuels

Knowledge of the genomics of microorganisms is being used to find better ways to harness biofuels from algae and cyanobacteria. The primary sources of fuel today are coal, oil, wood, and other plant products, such as ethanol. Although plants are renewable resources, there is still a need to find more alternative renewable sources of energy to meet our population's energy demands. The microbial world is one of

the largest resources for genes that encode new enzymes and produce new organic compounds, and it remains largely untapped. Microorganisms are used to create products, such as enzymes that are used in research, antibiotics, and other antimicrobial mechanisms. Microbial genomics is helping to develop diagnostic tools, improved vaccines, new disease treatments, and advanced environmental cleanup techniques.

Mitochondrial Genomics

Mitochondria are intracellular organelles that contain their own DNA. Mitochondrial DNA mutates at a rapid rate and scientists often use it to study evolutionary relationships. Another feature that makes studying the mitochondrial genome interesting is that the mitochondrial DNA in most multicellular organisms passes from the mother during the fertilization process. For this reason, scientists often use mitochondrial genomics to trace genealogy.

Experts have used information and clues from DNA samples at crime scenes as evidence in court cases, and they have used genetic markers in forensic analysis. Genomic analysis has also become useful in this field. The first publication showcasing the first use of genomics in forensics came out in 2001. It was a collaborative attempt between academic research institutions and the FBI to solve the mysterious cases of anthrax communicated via the US Postal Service. Using microbial genomics, researchers determined that the culprit used a specific anthrax strain in all the mailings.

Genomics in Agriculture

Genomics can reduce the trials and failures involved in scientific research to a certain extent, which could improve agricultural crop yield quality and quantity. Linking traits to genes or gene signatures helps improve crop breeding to generate hybrids

with the most desirable qualities. Scientists use genomic data to identify desirable traits, and then transfer those traits to a different organism. Researchers are discovering how genomics can improve agricultural production's quality and quantity. For example, scientists could use desirable traits to create a useful product or enhance an existing product, such as making a droughtsensitive crop more tolerant of the dry season.

Genomics and Proteomics

Proteins are the final products of genes, which help perform the function that the gene encodes. Amino acids comprise proteins and play important roles in the cell. All enzymes (except ribozymes) are proteins that act as catalysts to affect the rate of reactions. Proteins are also regulatory molecules, and some are hormones. Transport proteins, such as hemoglobin, help transport oxygen to various organs. Antibodies that defend against foreign particles are also proteins. In the diseased state, protein function can be impaired because of changes at the genetic level or because of direct impact on a specific protein.

A **proteome** is the entire set of proteins that a cell type produces. We can study proteoms using the knowledge of genomes because genes code for mRNAs, and the mRNAs encode proteins. Although mRNA analysis is a step in the right direction, not all mRNAs are translated into proteins. **Proteomics** is the study of proteomes' function. Proteomics complements genomics and is useful when scientists want to test their hypotheses that they based on genes. Even though all multicellular organisms' cells have the same set of genes, the set of proteins produced in different tissues is different and dependent on gene expression. Thus, the genome is constant, but the proteome varies and is dynamic within an organism. In addition, RNAs can be alternately spliced (cut and pasted to create novel combinations and novel

proteins) and many proteins modify themselves after translation by processes such as proteolytic cleavage, phosphorylation, glycosylation, and ubiquitination. There are also protein-protein interactions, which complicate studying proteomes. Although the genome provides a blueprint, the final architecture depends on several factors that can change the progression of events that generate the proteome.

Metabolomics is related to genomics and proteomics. **Metabolomics** involves studying small molecule metabolites in an organism. The **metabolome** is the complete set of metabolites that are related to an organism's genetic makeup. Metabolomics offers an opportunity to compare genetic makeup and physical characteristics, as well as genetic makeup and environmental factors. The goal of metabolome research is to identify, quantify, and catalogue all the metabolites in living organisms' tissues and fluids.

Basic Techniques in Protein Analysis

The ultimate goal of proteomics is to identify or compare the proteins expressed from a given genome under specific conditions, study the interactions between the proteins, and use the information to predict cell behavior or develop drug targets. Just as scientists analyze the genome using the basic DNA sequencing technique, proteomics requires techniques for protein analysis. The basic technique for protein analysis, analogous to DNA sequencing, is mass spectrometry. Mass spectrometry identifies and determines a molecule's characteristics. Advances in spectrometry have allowed researchers to analyze very small protein samples. X-ray crystallography, for example, enables scientists to determine a protein crystal's three-dimensional structure at atomic resolution. Another protein imaging technique, nuclear magnetic resonance (NMR), uses atoms' magnetic properties to determine the protein's three-dimensional structure in aqueous solution. Scientists

have also used protein microarrays to study protein interactions. Large-scale adaptations of the basic two-hybrid screen (Figure 63) have provided the basis for protein microarrays. Scientists use computer software to analyze the vast amount of data for proteomic analysis.

Genomic- and proteomic-scale analyses are part of **systems biology**, which is the study of whole biological systems (genomes and proteomes) based on interactions within the system. The European Bioinformatics Institute and the Human Proteome Organization (HUPO) are developing and establishing effective tools to sort through the enormous pile of systems biology data. Because proteins are the direct products of genes and reflect activity at the genomic level, it is natural to use proteomes to compare the protein profiles of different cells to identify proteins and genes involved in disease processes. Most pharmaceutical drug trials target proteins. Researchers use information that they obtain from proteomics to identify novel drugs and to understand their mechanisms of action.

Scientists are challenged when implementing proteomic analysis because it is difficult to detect small protein quantities. Although mass spectrometry is good for detecting small protein amounts, variations in protein expression in diseased states can be difficult to discern. Proteins are naturally unstable molecules, which makes proteomic analysis much more difficult than genomic analysis.

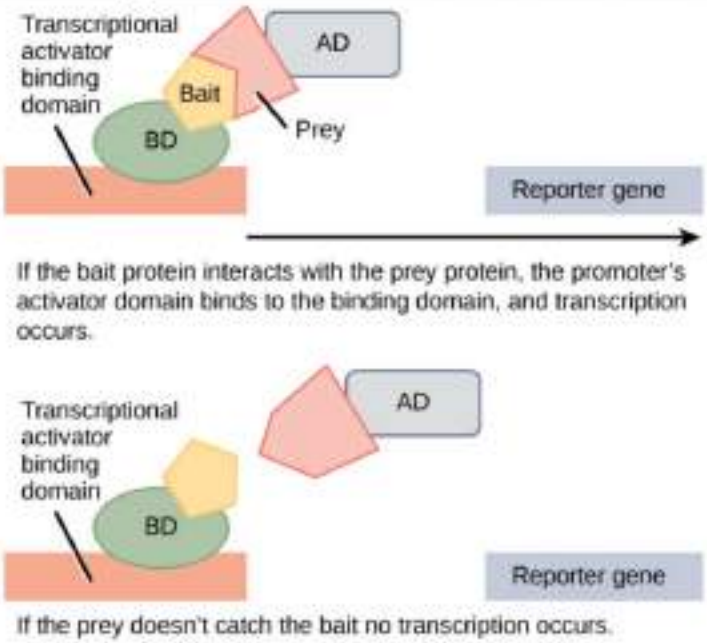


Figure 63: Scientists use two-hybrid screening to determine whether two proteins interact. In this method, a transcription factor splits into a DNA-binding domain (BD) and an activator domain (AD). The binding domain is able to bind the promoter in the activator domain's absence, but it does not turn on transcription. The bait protein attaches to the BD, and the prey protein attaches to the AD. Transcription occurs only if the prey "catches" the bait.

Cancer Proteomics

Researchers are studying patients' genomes and proteomes to understand the genetic basis of diseases. The most prominent disease researchers are studying with proteomic approaches is cancer. These approaches improve screening and early cancer detection. Researchers can identify proteins whose expression indicates the disease process. An individual protein is a **biomarker**, whereas a set of proteins with altered expression levels is a **protein signature**. For a biomarker or protein signature to be useful as a candidate for early cancer screening and detection, they must secrete in body fluids, such as sweat, blood, or urine, such that health professionals can perform large-scale screenings in a noninvasive fashion. The current problem with using biomarkers for early cancer detection is the high rate of false-negative results.

A **false negative** is an incorrect test result that should have been positive. In other words, many cancer cases go undetected, which makes biomarkers unreliable. Some examples of protein biomarkers in cancer detection are CA-125 for ovarian cancer and PSA for prostate cancer. Protein signatures may be more reliable than biomarkers to detect cancer cells. Researchers are also using proteomics to develop individualized treatment plans, which involves predicting whether or not an individual will respond to specific drugs and the side effects that the individual may experience. Researchers also use proteomics to predict the possibility of disease recurrence.

The National Cancer Institute has developed programs to improve cancer detection and treatment. The Clinical Proteomic Technologies for Cancer and the Early Detection Research Network are efforts to identify protein signatures specific to different cancer types. The Biomedical Proteomics Program identifies protein signatures and designs effective therapies for cancer patients.

APPLICATIONS OF BIOTECHNOLOGY

1. Biological Fuel Generation
2. Single-cell Protein
3. Sewage Treatment
4. Environmental Biotechnology
5. Medical Biotechnology
6 Agriculture and Forest Biotechnology
7. Food and Beverage Biotechnology
8. Safety in Biotechnology

1. Biological Fuel Generation

Organic residues to liquid fuels will become a more economically attractive consideration.

There are three main directions that can be followed to achieve biomass supplies: (1) cultivation of so-called energy crops, (2) harvesting of natural vegetation, and (3) utilization of agricultural and other organic wastes. The conversion of the resulting biomass to usable fuels can be accomplished by either biological or chemical means or by a combination of both. The two main end products that will be formed, will be either methane or ethanol although other products may arise depending on initial biomass and the processes utilized, for example, solid fuels, hydrogen, low-energy gases, methanol and longer-chain hydrocarbons.

Although biomass may ultimately only supply a relatively small amount of the world's energy requirements, it will nevertheless be of immense overall value. In some parts of the world, such as Brazil and countries of similar climatic conditions, biomass will surely attain wider exploitation and utilization. There may still be some disadvantages when comparing it with coal or oil, but the very fact that it is

renewable, and they are not must spur further research. In time, biomass will become much more easily available and economically useful as a source of energy for mankind.

2- Single-cell Protein

One of the biggest problems facing the world today is population growth, especially in developing nations. Conventional agriculture may not be able to provide sufficient supply of food and protein. Today, new agricultural practices are widespread, high-protein cereals have been developed, the cultivation of soybeans and groundnuts is ever expanding, and so on. The use of microbes as protein producers has also gained wide experimental success. This field of study has become known as **single-cell protein production (SCP)** and reflects the fact that most microorganisms used as producers grow as single or filamentous individuals rather than as complex multicellular organisms such as plants or animals.

There are many reasons microbes are the prime candidate for SCP production. Some of the reasons include: (1) microorganisms can grow at remarkably rapid rates under optimum conditions (some microbes can double their mass every 0.5 to 1 hour); (2) microorganisms are more easily modified genetically than plants and animals (they are more amenable to large-scale screening programs to select for higher growth rate, improved amino acid content, etc., and can be more easily subjected to gene transfer technology); (3) microorganisms have relatively high protein content, and the nutritional value of the protein is good; (4) microorganisms can be grown in vast numbers in relatively small continuous fermentation processes using a relatively small land area and are also independent of climate; and (5) microorganisms can grow on a wide range of raw materials, in particular low value wastes, and can also use plant-derived cellulose.

The acceptability of SCP, when presented as human food, depends not only on its safety and nutritional value, but also on other factors. People do not usually take to the idea of eating food derived from microbes. In many cultures, there are guidelines of what you can and cannot eat. Also, odor, color, taste, and texture need to be taken into consideration when dealing with people's desires. Thus, if SCP is to be used as direct food for humans, then the skills of the food technologist will be greatly challenged.

3- Sewage Treatment

Waste is, as any material or energy form that cannot be economically used, recovered or recycled at a given time and place. Growth in human populations has generally been matched by a greater formation of a wider range of waste products, many of which cause serious environmental pollution if allowed to accumulate in the ecosystem. In rural communities, recycling of human, animal, and vegetable waste has been practiced for centuries, providing in many cases valuable fertilizers or fuel. In urban communities where most of the deleterious wastes accumulate, efficient waste collection and specific treatment processes have been developed because it is impractical to discharge high volumes of waste into natural land and waters. The development of these practices in the last century was one of the main reasons for the spectacular improvement in health and well-being of humans.

Mainly by empirical means, a variety of biological treatment systems have been developed, ranging from cesspits, septic tanks, and sewage farms to gravel beds, percolating filters, and activated sludge processes coupled with anaerobic digestion. The primary aim of all these systems or biotreaters is to alleviate health hazards and to reduce the amount of oxidizable organic compounds and thus produce a final effluent or outflow, which can be discharged into the natural environment without

producing any adverse effect. Bioreactors rely on the metabolic versatility of mixed microbial populations for their efficiency. The fundamental feature of biotreaters is that they should contain a range of microorganisms with the overall metabolic capacity to degrade any compound entering the system. Controlled use of microorganisms has lead to the virtual elimination of such water-borne diseases as typhoid, cholera, and dysentery in industrialized communities.

4- Environmental Biotechnology

In industrialized countries, there is increasing public concern over the impact of human activities on the environment and the legacy we leave for future generations. Attention is being given to minimizing environmental damage, and to cleaning up past environmental damage, while trying to ensure that the standard of living to which we have been accustomed is maintained.

As an example, in the United States, $100 billion worth of crops are destroyed annually by soil-dwelling nematodes damaging crop plants. Chemical nematicides are the only current option for crop protection, but these are among the most toxic and environmentally damaging pesticides in widespread use.

It is obviously far better not to pollute the environment in the first place, than have to develop ingenious ways of cleaning up the environment. For this reason, many agricultural industries are investigating the potential of plants and microorganisms to provide cleaner processes and products.

The benefits of this research include:

- An increase in the productivity of crops, without an increase in the dependency on environmentally damaging agrochemicals.
- As a result of increased productivity, a reduced pressure to exploit the remaining uncultivated habitats.

- As a result of increased productivity, a reduction in energy inputs (mostly from reduced agrochemical manufacture).
- The creation of alternative, renewable, sources of energy (e.g., biodiesel).
- The creation of new more environment-friendly raw materials for industry (e.g., biodegradable plastics from plant starches, or high-value speciality chemicals).
- As a result of the development of genetically modified crops (if properly used), a reduction in the amount of agrochemical (e.g., pesticides and herbicides) released into the environment.

- **Environmental Concerns**

Many environmental groups, and some biotechnologists, have concerns about the use of genetically-modified crops with respect to their effect on the environment. They warn that the environmental impact of their use might not become apparent until after their large-scale commercial use. These concerns include:

Herbicide use. The use of herbicide-resistant crops may increase rather than decrease herbicide use.

Genetic pollution and superweeds. Genes that have been copied from one species and inserted into another might 'escape' and spread to other organisms, thus causing 'genetic pollution.' For example, herbicide-resistant crops might cross-breed with related weeds and produce herbicide-resistant 'superweeds.' **Antibiotic resistance.** Genes that code for antibiotic resistance are sometimes inserted into plant cells together with the 'useful' gene; these 'marker' genes enable scientists to select cells that have been successfully modified. Such

antibiotic-resistance markers in crops might spread to animals or humans, rendering medical or veterinary use of the antibiotic ineffective. **Unexpected effects.** Some genes that are inserted into genetically modified plants might be unstable (there is a high probability that they will be lost from the plant cells), or might show unexpected

effects because it is difficult to predict where in the plant genome the genes will insert.

Pest resistance. Genetically-modified crops that are designed to be resistant to a particular pest might have unintentional (and unpredictable) effects on harmless or beneficial organisms (e.g., ladybirds and bees).

Persistence and weediness. Genetically-modified plants might, intentionally or unintentionally, be more vigorous than their non-modified relatives. They can therefore effectively become 'weeds.' If plants are more 'persistent' (e.g., survive over winter better), they could rapidly dominate ecosystems at the expense of other plants. If they show 'weediness' characteristics, they could also spread to new habitats.

Damage to wildlife and biodiversity. Monocultures of pest-resistant crops may also harm beneficial insects and non-target organisms, with knock-on effects on the food chain.

- **Addressing Environmental Concerns**

Because it is difficult to assess the risks associated with the potential environmental impact of using genetically-modified crops, precautionary environmental legislation is in place in many countries. Before allowing experimental or commercial release of genetically-modified organisms, most countries use risk-assessment procedures that are based on estimating the impact and interactions of genetically-modified crops and their environment. The following counter-views might also be considered: Genetic Modification of Plants Achieves Essentially the Same Result as Conventional.

- **Plant Breeding**

The ability to move single genes, or sets of genes, from one species to another allows 'step changes' in genetic composition to be made quickly. However, in some cases,

the end product is effectively the same as might be achieved eventually by conventional breeding.

- **Wide-crossing**

'Wide-crossing' is regarded as a conventional breeding technique, but is used to create new varieties of crops between plants that would not normally interbreed. For example, some varieties of wheat that are used to make bread contain some genes from rye. This is the result of a plant-breeding experiment that introduced the rye genes to the wheat that confers disease resistance.

In Nature, Genes are Exchanged between Species

Bacteria that are not related can exchange genes quite readily (this is one mechanism for antibiotic-resistance spreading). *Agrobacterium tumefaciens* is a bacterium known as 'nature's own genetic engineer,' which routinely transfers some of its genes into plant cells.

- **Gene Instability in Conventional Crops**

The genes in new varieties of crops that have been developed by conventional breeding can also be unstable, particularly when they are produced as a result of wide-crosses. As with GM crops, selective breeding and variety evaluation 'weed out' undesirable gene combinations.

- **Conventional Breeding is not 'Natural' Either**

Plant breeders have for many years used techniques to introduce more variation than nature produces. Bombardment of a plant with chemicals or radiation causes mutations randomly throughout its genes. Among the plants that survive might be tens or thousands of mutations—perhaps including ones that plant breeders want.

- **Gene Transfer from Crops to their 'Wild' Relatives**

Little is known about the transfer of genes from any crop to the environment. 'Escaped' genes are most likely to spread to the crops' nearest relatives. Many, but

not all, crops are grown in environments in which their 'wild' relatives do not exist. For example, in Europe, genes are much less likely to transfer from genetically-modified potatoes than from those grown in Latin American countries where there are wild relatives of the potato. There is no evidence to suggest the genes in genetically-modified plants 'escape' any more easily than genes in other plants. Herbicide-tolerant crops—produced by conventional breeding—already exist, as does the possibility of gene transfer and 'superweeds.' To date this has not been a problem.

- **Novel, Sustainable, Agricultural Practices Needed**

There are concerns that pest-resistant plants will remove important food sources for wildlife and affect non-target organisms. The ecological impact of such crops must be evaluated, particularly for large-scale use. But we must also remember that many pesticides currently in use are not highly selective for pests, harming non-target organisms. Agriculture, as currently practiced, is the apparent cause of some disastrous effects on wildlife biodiversity—we must look for new options.

Application of biotechnology that we discussed in biological fuel generation, single-cell protein, single-cell oils, sewage treatment, etc., can also be covered under environmental biotechnology and can be used for the above purposes while cleaning the environment using biotechnological methods.

5- Medical Biotechnology

The use of biotechnology in medicine is growing rapidly and is providing us with opportunities to develop new, more effective drugs and other therapeutics. Studying the genetics of humans is allowing us to understand what happens when genes go wrong in inherited diseases, or cancers, and to start to develop new therapies that treat the genetic cause, not the symptoms. By studying the genetics of viruses, fungi

or bacteria, we can understand how they cause disease and develop better drugs and antibiotics that target them more specifically.

- **Safer, Cheaper Medicines from Biotechnology**

Many medicines have traditionally been extracted from plants, microorganisms, or animals. However, in many cases, it is not possible to produce the medicines in large enough amounts, safely, or cheaply enough to treat large numbers of people. Valuable drugs may be available only from endangered plant species (e.g., the anti-cancer drug tactual from the Pacific yew), or produced naturally in such small amounts that extraction is just too expensive. Biotechnology offers ways to produce valuable medicines in larger amounts by 'cell culture,' or using genetic modification to increase the amounts produced, or to make entirely new medicines.

- **Medicines from 'Cultured' Cells**

Many different types of plant, animal, and human cells can be grown 'in culture,' that is, as a suspension of cells in a large fermentor vessel of liquid that contains all of the nutrients they need, just as though they were yeasts or other microorganisms. These cells can be treated in various ways to make them produce large amounts of valuable compounds: microorganisms grown in this way can produce antibiotics. Plant cells grown in this way have been used to produce the anti-leukemia drugs vinblastine and vincristine: these are natural products of the pink periwinkle. Sedatives and heart drugs (e.g., digoxin from foxglove) have also been produced in plant culture.

- **Genetic Modification for Medicine Production**

Organisms can be genetically modified either to increase the amount of medically useful products, such as antibiotics that they produce, or to produce entirely new medicines. Microorganisms and plants are both being genetically modified to

produce large amounts of medicines safely and cheaply: genetically-modified microorganisms containing the gene for human insulin produce a safe and plentiful supply of insulin for diabetics; while genetically-modified plants are starting to be used to produce vaccines for a variety of human and animal diseases.

Research into finding new antibiotics is probably more urgent now than at any time since the discovery of penicillin some 50 years ago. Why? Many antibiotics that were once highly effective are now proving to be useless as the disease-causing microorganisms develop resistance. A new approach is being used to try to make entirely new antibiotics. A soil bacterium—*streptomyces*—is a source of antibiotics, immunosuppressants, anti-cancer and anti-parasitic agents, and natural herbicides. Many of these valuable products are made by the bacterium from smaller building blocks according to the 'blueprint' determined by the bacterium's genes. It is possible to generate new compounds by changing the blueprint—switching some of the genes, and thus some of the units, around. A whole range of new compounds can be made using this approach, potentially leading to new antibiotics. This area of research is extremely important, considering the rising number of diseases caused by antibiotic-resistant microorganisms.

- **Applications of Biotechnology in Human Health**

✓ ***Recombinant DNA Technology***

Combining DNA through natural sexual reproduction can occur only between individuals of the same species. Since 1972 technology has, however, been available that allows the identification of genes for specific, desirable traits and the transfer of these, often using a virus as the vector, into another organism. Comparable to a word-processor's 'cut-and-paste', this process is called recombinant DNA technology or gene splicing. Virtually any desirable trait found in nature can, in principle, be transferred into any chosen organism. An organism modified by gene splicing is

called transgenic or genetically modified (GM). Specific applications of this type of genetic engineering are rapidly increasing in number - in the production of pharmaceuticals, gene therapy, development of transgenic plants and animals, and in several other fields.

✓ ***Pharmaceutical Production***

The first major healthcare application of recombinant technology was in the production of human insulin, a hormone substantially involved in the regulation of metabolism, particularly

5 Proteus – in Greek mythology a god who knew all things past, present, and future but disliked telling what he knew. From his power of assuming whatever shape he pleased, Proteus came to be regarded as a symbol of the original matter from which all is created of carbohydrates and fats, and the relative lack of which leads to the clinical condition called diabetes mellitus. Insulin is a relatively small protein consisting of 51 amino acids.

While the bovine or porcine insulin that had been used to treat human diabetes since the 1920s had become increasingly pure, side effects did occur due to its originating from a different species. In 1978, however, scientists succeeded in inserting the gene for human insulin into an E. coli bacterium. Once inside the bacterial cell, the gene could turn on its bacterial host's protein making machine to make – human insulin. Bacterial cells divide rapidly to make billions of copies of themselves, each modified bacterium carrying in its DNA an accurate replica of the gene for insulin production. Thus, given the necessary environmental factors, the bacteria would produce significant quantities of insulin, which can then be extracted from the 'soup' in which the process takes place and purified for use in humans. Today, most commercially available insulin is produced in this manner, using e.g. yeast cells as hosts.

A perhaps more famous example is recombinant erythropoietin, a hormone that regulates the production of red blood cells. The clinical conditions for which erythropoietin is indicated are relatively rare, but the bio-engineered product has gained enormous popularity in professional sports – as EPO – because it enables athletes to add 15-20 per cent to their oxygen carrying capacity.

Using micro-organisms or human cell cultures, similarly modified, in the production of highly complex molecules which would otherwise be impossible, or extremely difficult, to synthesise, is now employed extensively by the pharmaceutical industry. Increasingly, higher animals - "bioreactors" – modified by recombinant technology and able to express high value pharmaceutical proteins in their milk are also gaining use in reducing the cost of creating and producing new medical products.

- ✓ ***Vaccines; Recombinant Technology and the Immune System***

A vaccine is an antigen, e.g. the surface proteins of a pathogenic micro-organism. By exposing the immune system to an antigen previously 'unknown' to it, it primes the system so that on later contact with the antigen, a swift and effective defence will be mounted to prevent disease. The substances involved in this defence are called antibodies, proteins specific to, and able to deactivate the germs that carry, the particular antigen 'remembered' from previous contact, e.g. from vaccination. Immunological memory, including the ability to produce specific antibodies, is held by specialised white blood cells, making use of their 'cell factory' as described above. Obviously, an antigen used as a vaccine should be unable to cause disease, or at the least be much less a threat than the organism against which it is intended to protect. The classic example is Jenner's use 200 years ago of cowpox (vaccinia)6 virus to immunize his son. While cowpox virus is almost a-pathogenic to humans, it has antigenic characteristics akin to those of the human smallpox virus – a close 'relative' – or close enough to induce an immune response sufficient to fight off

'real' smallpox. Immunisation is a cornerstone of preventive medicine, having provided some of the most cost-effective health interventions known.

Traditionally, vaccines are live attenuated (weakened virus or bacteria) or inactivated; the latter either whole, killed micro-organisms or e.g. selected cell surface proteins. While technological limitations remain and, for example, an effective AIDS/HIV vaccine has not yet been found, recombinant technology constitutes a powerful tool for the production of purer and safer vaccines. For example, the insertion of a hepatitis B virus gene into the genome of a yeast cell allows the production of pure hepatitis B surface antigen - a very effective vaccine, biologically equivalent of an inactivated vaccine. A live attenuated typhoid vaccine is now being produced from a *Salmonella typhi* bacterium cell line modified by recombinant technology so as not to cause typhoid. Several new vaccines using genetically weakened

Separately, recombinant technology is now being used to modify plants, rather than animal cell lines or micro-organisms, to produce vaccines. Likely to gain increased use in the future, this will enable many vaccines to be made for oral administration, thus overcoming many vaccine logistics constraints and the need for medically qualified or veterinary personnel and other costly elements currently necessary to carry out effective immunisations. The first potato-produced, edible hepatitis B vaccine is in clinical trial.

In addition to vaccines to *prevent* against micro-organisms, others – so-called therapeutic vaccines - based on combining immune pathology and genetic modification may soon revolutionise the *treatment* of many diseases – infectious as well as non-infectious. Some of these will stimulate an impaired immune response in an individual who is already infected with that organism and has mounted an inadequate immune response to that organism. The aim of administering a

therapeutic vaccine may be to increase the individual's immunity to an organism that, for instance, is unable to provoke an appropriate response on its own. A vaccine against Helicobactor pylori, the causative agent of duodenal ulcers is being tested. Other vaccine approaches under development modulate the immune response in rheumatoid arthritis and related disorders, the pathological mechanisms of which involve an inappropriate, so-called autoimmune process. Similarly, vaccines are being developed for use in the treatment of diseases, such as asthma, hypertension, atherosclerosis, Alzheimer's disease and others, in which so-called endogenous7 substances, are known to play a role. Also, and perhaps at an even more advanced stage, there are vaccines against specific cancers, e.g. melanoma, breast cancer, colon cancer8, or even one that may offer more universal protection against cancer.9 Not related to vaccines, but nevertheless at the epistemological intersection of immunology and recombinant technology, attempts are underway to modify the coding – by cut-and-paste recombinant technology – for the so-called immunomodulators. These are naturally occurring molecules (cytokines, interleukins, interferons) with broader, regulatory effects on the immune system, as well as on several other biological functions, such as wound healing, nerve cell repair, blood cell formation. While the use of interferon – as a drug - in multiple sclerosis has been the topic of a recent debate, the ability to adjust 'own' production of these modulators may have important applications in a majority of the diseases currently plagueing mankind.

- ✓ ***Monoclonal antibodies***

While vaccines are antigens which, when inoculated, cause the immune system to produce antibodies, recombinant technology is being used, as well, to produce antibodies directly. In this variation on the immune/genetics theme, single cell lines, i.e. cloned, wholly identical, specialised cells that can be grown indefinitely are used

to produce antibodies of singular specificity - monoclonal antibodies. These are used in a number of diagnostic applications, as well as to prevent acute transplant rejection, and treat leukaemias and lymphomas. Some show promise against auto-immune diseases.

✓ ***Gene Therapies***

While the above applications mostly rely on using modified organisms or cell lines to produce substances *in vitro* that can then be used to treat or prevent human disease, gene therapy is distinctly different in that it essentially modifies the patient's own genetic setup. In other words, while the aim remains the manipulation of a specific gene into a designated host cell, the 'host' is a 'population' of cells in situ in the human body. In contrast to the above technologies, gene therapy takes place *in vivo* Technical details differ, but gene therapy essentially makes use of an approach similar to recombinant technology. An isolated gene encoding for the desired characteristic is spliced into the genome of a virus11, often itself modified so as not to cause disease. Infecting the host organism, the virus introduces the gene into the target cells to 'appropriate' the cells' protein-making apparatus. Gene treatment is likely to involve one of the following:

• **Gene replacement,** a substitution of a non-active or defective gene by a "new" (or additional), functional copy of the gene, to restore the production of a required protein. This technique is used in e.g. the treatment of cystic fibrosis and certain cancers;

• **Gene addition,** the insertion into the cell of a *new* gene, to enable the production of a protein not normally expressed by that cell. For example, the code for a stimulatory protein may be inserted to enhance an immune response to cancer cells;

• Gene control, the alteration of expression of a gene used, for example, to suppress a mutated onco-gene in tumour cells so as to prevent specific protein production.

Gene therapy was first used in 1990, for an enzyme deficiency. Since then, more than 100 clinical gene-therapy trials have been initiated world-wide. Most of the trials have been for the treatment of tumours (predominantly malignant melanoma and haematological disorders), but there have also been trials of gene therapies for genetic disorders, AIDS, and cardiovascular disease. While many technical problems are yet unsolved, in relation to vector design as well as to clinical safety and efficacy, gene therapy appears likely to become an important part of the armoury with which disease will be fought in the future.

- ***Other Medical Biotechnology Applications***

Stem cell research and cloning share technological approaches and are occasionally combined with recombinant technologies. However, rather than the 'cut-and-paste' approach to DNA in recombinant technology, the central premise of stem cell and cloning is to preserve the entirety of the genome and guide its ability to express itself for novel therapeutic applications.

✓ ***Stem Cell Research and its Potential***

Upon fertilization, an egg cell initially starts dividing into undifferentiated cells from which, later, cells of increasing specialization develop and from which eventually the highly differentiated cells in tissues of different organs stem. In human embryos, the potential for giving rise to cells of any specialisation is held only by very early, primitive, so-called *totipotent* stem cells, at the most up to the 16-cell stage. Identical twins (triplets etc.) originate from totipotent cells, i.e., the result of a cleavage of the embryo within a few days after fertilization.

At the next stage of development, the now *pluripotent* stem cells have already acquired some degree of specialization. While they are no longer individually able to give rise to a foetus, they are still able to differentiate into any cells of an adult human being. *Multipotent* stem cells can be derived from foetuses or umbilical cord

blood, and are even present throughout life, although in progressively decreasing numbers in adults. Unless they are 'reprogrammed', the latter cells are probably only able to develop into specialized tissues or organs. Common to stem cells is their ability - under given circumstances - to multiply almost indefinitely and be stimulated to grow into a variety of specialized tissues, opening up vast possibilities of tissue repair.

Much of the controversy over stem cell research relates to the ethics of using cells deriving from aborted fetuses, seen by many as a violation of the respect for human life. In recognition of this, the debate has partly centered on the possibility of allowing stem cell research to be carried out on early embryos no longer needed for infertility treatment ("spare embryos") or resulting from in vitro fertilization specifically for research. However, ethical concerns also arise from the potential of creating stem cells by cell nuclear replacement.

This technique involves removing the nucleus, i.e., the DNA, of an egg and replacing it with the nucleus of a cell from a given individual. This would enable the cultivation of pluripotent cells genetically almost identical to the person from which the nucleus was derived. Such cells would therefore not evoke an immune rejection, and transplant medicine would offer entirely new therapies. The problem, in moral terms, with nuclear transfer is its likeness - technically - to cloning, the creation of a true copy of an existing individual. However, while cloning and this particular pursuit of stem cell research largely share the technique of nuclear replacement, they differ significantly in that the latter involves the extraction of stem cells for the purpose of developing the tissue of a single organ - the heart, nerve cells etc.

The potential scope of stem cell research and derived applications is enormous. Improved transplantation therapy with tissue grown from stem cells in a laboratory would open the possibility of renewing heart muscle in congestive heart failure;

replacing blood-forming stem cells to produce healthy red and white blood cells to treat e.g. AIDS and leukaemias; relining blood vessels with new cells as treatment for atherosclerosis, angina, or stroke; restoring islet cells in the pancreas to produce natural insulin in diabetics; or renewing of nerve cells in patients with Parkinson's disease or paralysis. Stem cell therapy may also bring a host of rare congenital disorders within therapeutic reach.

✓ ***Cloning***

Human cloning has become a highly emotive issue. However, unsensational, and far from uncommon in nature, a clone is essentially the result of asexual reproduction, leaving clones with no choice but to accept a genome identical to that of their ancestor. Microbes reproduce by cloning; the chrysanthemum plants available at the local supermarket are clones of a long dead plant, as are the high-yielding vines in a Bordeaux vineyard. And one of a pair of identical twins is a clone of the other.

Cloning in modern biotechnology is based on cell nucleus transfer, and Dolly, the first mammal to be cloned, is the result of a transfer of the nucleus of an udder cell to an enucleated egg cell. Following this, the egg was implanted in the mother's uterus and went through a normal gestation. Contrary to public expectation, Dolly may not have made the cloning of a human being any likelier to happen; it simply may not be possible - other than in fiction. For while the principle would be the same as in sheep, 'switching' the genetic complement in the nucleus of, say, a skin cell from performing its rather specialized functions to taking on the highly complex role of orchestrating embryonic differentiation and development may not be feasible in some species, given a very limited 'window of opportunity'. Cloning a mouse, a mammal far better known as a laboratory animal than sheep, was tried unsuccessfully for a long time12 and, after all, Dolly was the only success among about 300 attempts.

Even if human cloning were possible, its appeal may well be more fictional than real - partly a result of literary and cinematic hype. Aside from 'vanity cloning', a real demand for which remains dubious, cloning of humans may be of little value other than to those who are childless as a result of genetic disease. With a success rate of less than one per cent, however, this option hardly looks interesting. Add the many unknown factors related to the resulting child's genetic predisposition and the attractiveness of human cloning remains dubious. Thus, with no demonstrable benefits - and few supporters - prohibiting human reproductive cloning would appear to be straightforward.

Emphasizing this point, the cloning of mammals has no value from the point of view of breeding of farm animals; for that, it remains far too risky and costly. Most, if not all, of its attraction derives from its potential in pharmaceutical production. Of particular allure is the potential of having animals' express proteins of therapeutic value in their milk. Interestingly though, this will be achieved through recombinant technology, i.e., insertion of the appropriate gene, as earlier described, rather than of cloning *per se*. In the context, however, cloning, is intended to enable the breeding of animals with a genetic setup that facilitates, or impedes least, the production of the required pharmaceuticals.

6-Agriculture and Forest Biotechnology

Genetic engineering can be used to modify the genetic compositions of plants, animals, and microorganisms. The number of genes that have been isolated and are available for transfer is growing daily. Currently, the technology is used primarily to modify crops, although a number of other applications are in the wings.

Like other products, genetically engineered products undergo a period of research and development before they are ready for commercial release. Many products never emerge from the research and development pipeline. While this is true for almost

any technology, genetic engineering is turning out to be more difficult and more expensive than early proponents expected. Although in the early 1980s biotechnology was touted as a miracle technology that was going to usher in a new era of agricultural abundance with minimal harm to the environment, the initial set of products have proven modest. Some of the most important commercial applications of biotechnology include:

- Crops that are resistant to pests or herbicides, which will allow reduced or more selective use of agrochemicals.
- Crops that are better able to tolerate hostile environmental conditions such as frost or drought.
- New and more efficient crop-breeding systems that allow production of hybrids that cannot be produced by conventional plant breeding.
- New crop varieties that are qualitatively different (e.g., plants that produce seeds or tubers that have an altered starch composition, and thus represent new valuable products).
- New crop varieties that have different growth characteristics (e.g., altered flowering time or rate of growth) to increase yield.

These tools fall into two major categories: improving crop performance in the field and developing new products with enhanced values.

- **Genetic Improvement through Traditional Crop Breeding**

The cultivation and improvement of food crops began centuries ago, when seeds from the most successful plants of one season were used to sow the next year's crop. Many of our modern cereals bear little resemblance to their 'wild' grass ancestors. Through breeding and selection, the characteristics of these crops have been greatly altered. Left to right: pearl millet, barley, wheat, maize, rice, and sorghum.

Improved crop varieties have been developed by selection and breeding, and few of today's crops resemble their wild ancestors. Some crops were developed by these conventional plant-breeding processes to be disease resistant or to grow faster, for example; others to make the edible parts of the plant larger or tastier. However, developing new varieties of crops by traditional methods is not only slow but also limits plant breeders to transferring genes between crops that normally crossfertilize.

- **Plant Genetics**

Over the past few years, scientists have been able to identify the individual genes that determine particular characteristics in a plant. Just as importantly, we can now transfer these genes between plants, which means that novel features can be introduced into crop species. It is unlikely that biotechnology will solve every agronomic problem, but it can offer farmers new options.

- **Plant Physiology and Biochemistry**

Studies of plant physiology, and biochemistry, have given us an understanding of how plants sense their environment, and respond by germinating, growing, flowering, or setting seed. All these processes are determined by their genes. How they photosynthesize, and make proteins, oils, or carbohydrates is now well understood. How they recognize insect pests or disease-causing fungi, bacteria or viruses (pathogens), and mount their own defense responses is also much better understood.

- **Genetic Modification of Crop Plants**

The discovery that genes are naturally transferred from the bacterial pathogen *agrobacterium tumefaciens* when it infects plants allowed a breakthrough in introducing new genes and desirable traits into crop plants. The subsequent development of other genetic modification methods means that it is now quite

straightforward to introduce a gene, or set of genes, into most of the world's major crops. Advanced breeding program means that genes introduced by genetic modification can often be moved into 'elite varieties' (highly productive and commercially successful varieties) within economically viable timeframes. A better understanding of how genes work is allowing the introduced genes to be controlled in quite sophisticated ways: switched 'on' or 'off' at a particular time during plant development or in response to environmental signals. For example, research is underway to make plants switch from using their energy supplies for 'growing tall' to storing energy in the form of starch and sugars. Other studies are aimed at getting plants to switch on the genes for defence mechanisms in response to attack by pests or diseases, or at stages of growth when they are most likely to be susceptible to attack.

- **Engineered Crops**

The most widespread application of genetic engineering in agriculture by far is in engineered crops. Thousands of such products have been field-tested and over a dozen have been approved for commercial use. The traits most introduced into crops are herbicide tolerance, insect tolerance, and virus tolerance.

- **Herbicide Tolerance**

Herbicide tolerance allows crops to withstand otherwise lethal doses of herbicides, which are chemicals that kill plants. Some herbicides kill virtually all plants and cannot be used on crops. By offering crops tolerant to herbicides, chemical companies can expand the market for their products. And indeed, the major developers of herbicide-tolerant plants are companies with herbicides to sell. The current set of commercially available herbicide-tolerant crops is tolerant to three herbicides based on three active ingredients: bromoxynil, glyphosate, and glufosinate.

- **Insect Tolerance**

All of the commercially available insect-tolerant plants contain a version of the toxin *bacillus thuringiensis* (Bt), which is found in nature in soil bacteria. Bt toxins are highly effective for many pest organisms, such as beetles and moth larva, but not toxic to mammals and most other non-target organisms. A major concern among farmers and environmentalists is that wide use of Bt crops will lead to the rapid development (over the course of perhaps as few as three to five years) of resistance to the toxin. If resistance develops, the Bt toxin will be as useless as a pesticide. In this case, the environmental benefits of the product will be short lived.

Loss of Bt will affect those who currently use the engineered Bt crops, but also many other farmers who use Bt in its natural bacterial form, usually as a spray. These other farmers include those who grow food organically and those who use Bt as part of integrated pest management (IPM). Natural Bt sprays are a valuable mode of pest control for these farmers. Organic farmers and others who rely on Bt question whether the companies who sell the Bt crops have the right to use up this resource guided only by commercial calculations.

- **Virus Tolerance**

The third major application of biotechnology to crops is virus tolerance. These crops contain a gene taken from a virus. By a process that is not well understood, plants that produce certain viral proteins are able to fend off infections by the viruses from which the proteins were taken. Two virus-tolerant crops are currently approved for commercial use: papaya and squash. The squash, which is resistant to two viruses, is currently off the market. Although it is difficult to get information on why products are not on the market, it is possible that the squash did not perform well enough in the field to capture market share.

- **Other Engineered Products**

Many other genetically engineered products have been envisioned, but only a few have come to market so far.

Several tomatoes engineered to delay ripening have been approved for commercial use. In some cases, delayed ripening just prolongs shelf life. But, for the Flavr SavrTM, the objective was to increase the time on the vine without softening, producing a transportable, tasty winter tomato. After a highly publicized rollout, the Flavr SavrTM too is off the market. The problem appears to have been with transportability rather than taste.

On the livestock side, a drug has been produced for dairy cows—Bovine Growth Hormone (BGH) or Bovine Somatotropin (BST)—by engineering a bacterium to contain the gene for the hormone. The drug is administered to cows to increase milk production, despite the chronic oversupply of milk in the United States. A highly controversial product when it was first introduced, BGH is currently used on about 10 percent of U.S. dairy herds.

An interesting product not related to agriculture is a rabies vaccine intended for use on wild raccoons. In this case, genetic engineering was used to construct a 'hybrid' virus made up of a component of the rabies virus inserted into an unrelated 'carrier' virus. The resulting virus confers immunity to rabies but poses no danger of causing the disease. Baits laced with the vaccine have been distributed in many parts of the eastern United States in attempts to combat rabies in wild raccoon populations. The vaccine has been approved by the USDA, despite suspicions that it has been only marginally, if at all, effective. Early studies on efficacy failed to demonstrate that the product could control rabies in wild raccoon populations. Data from more recent studies are being withheld from the public as confidential business information.

- **Agricultural Biotechnology Research Projects**

The following lists some of the organisms being engineered by agricultural researchers and indicates which products are commercially available and which are not.

- **Genetically Engineered Livestock and Poultry**

✓ **Animals Engineered for Leaner Meat**

No livestock engineered for leaner meat is currently near commercialization. Research done early in the 1980s to genetically engineer leaner pigs failed because of unacceptable side effects, including low fertility, arthritis, and impaired immune systems. Some low level of research activity may still be underway.

✓ **Animals Engineered as Drug-Production Facilities**

Status: Goats and sheep have been engineered to secrete bioactive molecules into their blood, urine, or milk. Companies are in the process of developing commercial enterprises based on these animals. So far, none of the drugs is on the market. It is likely that producers will want to slaughter the animals for food after they are no longer useful for drug production.

✓ **Animals Engineered as Sources of Transplant Organs**

Status: Commercial entities are engineering pigs so that their organs will not be rejected by human transplant recipients. So far, the organs are not commercially available. It is likely that producers will want to use the carcasses of donor pigs as food.

✓ **Animals Engineered for Disease Resistance**

Status: Chickens and turkeys have been engineered to resist avian diseases. None have been commercialized.

✓ **Genetically-Engineered Fish and Shellfish**

Status: Fish and shellfish have been engineered to cause changes in hormones that accelerate growth in several laboratories. So far, none have been commercialized in the United States.

✓ **Genetically-Engineered Plants Eaten Whole as Food**

Status: Many plants have been commercialized, including tomatoes and squash and commodity crops such as corn and soybeans. Most have been engineered for one of three traits: herbicide tolerance, insect resistance, or virus tolerance.

✓ **Genetically-Engineered Fiber Plants**

Status: Genetically-engineered cotton has been approved for commercial use.

✓ **Engineered Insects Used in Agricultural Systems**

Status: No engineered insects have been approved for commercial use. An engineered predatory mite has been field tested in Florida. Researchers have produced honeybees and other beneficial insects engineered to tolerate pesticides.

Engineered Microorganisms Used as Pesticides

Status: Several bacteria engineered to enhance their ability to kill or repel pests have been approved for commercial use. These products are used as pesticides in agricultural fields and gardens.

✓ **Food Processing Aids Made from Engineered Bacteria**

Status: Bacteria have been genetically engineered to produce rennet, an enzyme important in making cheese. Genetically engineered rennet (chymosin) is approved for commercial use and widely used by US cheese processors.

7. Food and Beverage Biotechnology

✓ Application of Modern Biotechnology in the Food Industry

The techniques of modern biotechnology are becoming an increasingly important part of the overall effort to improve methods of food production and to increase the variety and quality of foods. Modern biotechnology has potential applications in the production of food, food processing, and also in the assurance of food quality and safety.

Although many genetically modified food crops are under development, very few have so far reached the supermarket shelf. The first genetically modified whole food to reach the shops was the Flavr SavrTM tomato in the USA, and the first product in the UK was tomato puree. The recent introduction of genetically modified soybeans has caused controversy. Up to 60% of processed foods contain soy or soy derivatives, such as starch, protein, or oils.

Modern biotechnology is also being used to improve microorganisms, processing aids (e.g., enzymes) or ingredients for use in the production and processing of food. The enzyme chymosin, which is used in the manufacturing of 'vegetarian cheese,' is an example of this type of genetically modified product.

8. Safety in Biotechnology

Based on international guidelines, governments in each country frame rules to ensure not only the safety of the product but also for those who conduct the genetic engineering experiments. When the first rDNA experiments were carried out, scientists themselves imposed a moratorium on further genetic experiments until proper guidelines were structured. In 1977, the Asilomar Conference formulated some guidelines, and the major points are the following:

- Recombinant DNA experiments should be carried out in laminar flow-chamber kept in a clean room, so that the recombinant organisms do not escape.

- The host organism used to carry out the gene cloning or the cloning of rDNA should be specially made for such experiments. Even if the organism escapes the laboratory, it will not survive, and the recombinant DNA molecule will be lost. There are special safe vectors that can be used for rDNA experiments without fear. There is also non-pathogenic *e. coli* developed to use in rDNA experiments as host cells.

The products, in most cases, of modern biotechnology will be therapeutic proteins such as vaccines, hormones, and enzymes. The proteins that are to be used as drugs should meet all the safety requirements and regulations imposed by the regulatory authorities. They are very strict about the data regarding the experiments on animal models, terminal patience, non-terminal patience, and independent clinical trials by separate agencies. If a transgenic organism is released into the environment, in addition to the ordinary safety norms it should also monitor its environmental impact. The clinical trials of drugs produced by rDNA methods must be carried out at least in three phases.

Phase I. In the first phase of study, the side effects and dose of tolerance by the patients is systematically carried out on selected patients.

Phase II. In the second phase, collecting data from pharmacological, pharmacokinetic, metabolic, and toxicological studies on a selected number of patients optimizes the use of the drug.

Phase III. In the last phase, the studies focus on the safety aspects of the drugs. During these studies the harmful effects of the drug, if any, are monitored in addition to the effectiveness as a drug. In dosage range, the interactions with other drugs are also investigated.

A new biotechnology product, whether it is a drug or a food material, has to undergo trial studies, and only after the regulatory agencies are fully satisfied by the data, will they give permission or license it for mass production and marketing.

TERMINOLOGIES RELATED TO BIOTECHNOLOGY

- **Antibiotic resistance** ability of an organism to be unaffected by an antibiotic's actions.
- **Biomarker** individual protein that is uniquely produced in a diseased state.
- **Biotechnology** use of biological agents for technological advancement.
- **cDNA library** collection of cloned cDNA sequences.
- **Cellular cloning** production of identical cell populations by binary fission.
- **Chain termination method** method of DNA sequencing using labeled dideoxy nucleotides to terminate DNA replication; it is also called the dideoxy method or the Sanger method.
- **Clone** exact replica.
- **Contig** larger sequence of DNA assembled from overlapping shorter sequences.
- **Cytogenetic mapping** technique that uses a microscope to create a map from stained chromosomes.
- **Deoxynucleotide** individual DNA monomer (single unit)
- **Dideoxynucleotide** individual DNA monomer that is missing a hydroxyl group (–OH).
- **DNA microarray** method to detect gene expression by analyzing many DNA fragments that are fixed to a glass slide or a silicon chip to identify active genes and identify sequences.
- **Expressed sequence tag (EST)** short STS that is identified with cDNA.
- **False negative** incorrect test result that should have been positive.

- **Foreign DNA** DNA that belongs to a different species or DNA that is artificially synthesized.
- **Gel electrophoresis** technique used to separate molecules on the basis of size using electric charge.
- **Gene targeting** method for altering the sequence of a specific gene by introducing the modified version on a vector.
- **Gene therapy** technique used to cure inheritable diseases by replacing mutant genes with good genes.
- **Genetic diagnosis** diagnosis of the potential for disease development by analyzing disease-causing genes.
- **Genetic engineering** alteration of the genetic makeup of an organism.
- **Genetic map** outline of genes and their location on a chromosome.
- **Genetic marker** gene or sequence on a chromosome with a known location that is associated with a specific trait.
- **Genetic recombination** DNA exchange between homologous chromosome pairs
- **Genetic testing** process of testing for the presence of disease-causing genes.
- **Genetically modified organism (GMO)** organism whose genome has been artificially changed
- **Genome annotation** process of attaching biological information to gene sequences
- **Genome mapping** process of finding the location of genes on each chromosome.
- **Genomic library** collection of cloned DNA which represents all of the sequences and fragments from a genome.

- **Genomics** study of entire genomes including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species.
- **Host DNA** DNA that is present in the genome of the organism of interest
- **Linkage analysis** procedure that analyzes recombining genes to determine if they are linked
- **Lysis buffer** solution to break the cell membrane and release cell contents
- **Metabolome** complete set of metabolites which are related to an organism's genetic makeup.
- **Metabolomics** study of small molecule metabolites in an organism.
- **Metagenomics** study of multiple species' collective genomes that grow and interact in an environmental niche.
- **Microsatellite polymorphism** variation between individuals in the sequence and number of microsatellite DNA repeats.
- **Model organism** species that researchers study and use as a model to understand the biological processes in other species represented by the model organism.
- **Molecular cloning** cloning of DNA fragments.
- **Multiple cloning site (MCS)** site that multiple restriction endonucleases can recognize.
- **Next-generation sequencing** group of automated techniques for rapid DNA sequencing.
- **Northern blotting** transfer of RNA from a gel to a nylon membrane
- **Pharmacogenomics** study of drug interactions with the genome or proteome;

also called toxicogenomic.

- **Physical map** representation of the physical distance between genes or genetic markers.
- **Polygenic** phenotypic characteristic caused by two or more genes.
- **Polymerase chain reaction (PCR)** technique to amplify DNA.
- **Probe** small DNA fragment to determine if the complementary sequence is present in a DNA sample.
- **Protease** enzyme that breaks down proteins.
- **Protein signature** set of uniquely expressed proteins in the diseased state.
- **Proteome** entire set of proteins that cell type produces.
- **Proteomics** study of proteomes' function.
- **Pure culture** growth of a single cell type in the laboratory.
- **Radiation hybrid mapping** information obtained by fragmenting the chromosome with x-rays.
- **Recombinant DNA** combining DNA fragments that molecular cloning generates that do not exist in nature, also a chimeric molecule.
- **Recombinant protein** a gene's protein product derived by molecular cloning.
- **Reproductive cloning** entire organism cloning.
- **Restriction endonuclease** enzyme that can recognize and cleave specific DNA sequences.
- **Restriction fragment length polymorphism (RFLP)** variation between individuals in the length of DNA fragments, which restriction endonucleases generate.

- **Reverse genetics** method of determining the gene's function by starting with the gene itself instead of starting with the gene product.
- **Reverse transcriptase PCR (RT-PCR)** PCR technique that involves converting RNA to DNA by reverse transcriptase.
- **Ribonuclease** enzyme that breaks down RNA.
- **Sequence mapping** mapping information obtained after DNA sequencing.
- **Shotgun sequencing** method used to sequence multiple DNA fragments to generate the sequence of a large piece of DNA.
- **Single nucleotide polymorphism (SNP)** variation between individuals in a single nucleotide.
- **Southern blotting** DNA transfer from a gel to a nylon membrane.
- **Systems biology** study of whole biological systems (genomes and proteomes) based on interactions within the system.
- **Ti plasmid** plasmid system derived from Agrobacterium tumifaciens that scientists have used to introduce foreign DNA into plant cells.
- **Transgenic** organism that receives DNA from a different species.
- **Variable number of tandem repeats (VNTRs)** variation in the number of tandem repeats between individuals in the population.
- **Whole-genome sequencing** process determines an entire genome's DNA sequence.

REFRENCES

Bolsover, S. R., Shephard, E. A., White, H. A., & Hyams, J. S. (2011). *Cell biology: a short course* (Vol. 1). John Wiley & Sons.

Murray, R.K., Granner, D.K., Mayes, P.A. and Rodwell, V.W. 2003. Harper's Illustrated Biochemistry. 26th Edition. McGraw-Hill.

Nair, A. (2007). *Introduction to biotechnology and genetic engineering*.

Shmaefsky, B. (2006). *Biotechnology 101* . Greenwood Press.

Thieman, W. J., & Palladino, M. A. (2012). *Introduction to biotechnology*. Benjamin Cummings.

Vats M.; Kumari M. & Bhandari S. (2017). Cell and Molecular Biology. Uttarakhand Open University, Haldwani, Nainital- 263139.