

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ
КАФЕДРА БИОЛОГИИ

МЕДИЦИНСКАЯ БИОЛОГИЯ

MEDICAL BIOLOGY

В двух частях

Часть 1

Рекомендовано учебно-методическим объединением
по высшему медицинскому, фармацевтическому образованию
в качестве учебно-методического пособия для студентов
учреждений высшего образования, обучающихся
по специальностям 7-07-0911-01 «Лечебное дело»,
7-07-0911-03 «Стоматология», 7-07-0912-01 «Фармация»

2-е издание, исправленное



Минск БГМУ 2025

УДК 57:61(075.8)

ББК 28.0я73

М42

А в т о р ы: ст. преп. В. В. Григорович; ассист. Ю. И. Корбут; ст. преп. Е. А. Черноус; канд. мед. наук, доц. В. Э. Бутвиловский; канд. мед. наук, доц. В. В. Давыдов

Р е ц е н з е н т ы: д-р мед. наук, проф., зав. каф. медицинской биологии и общей генетики Витебского государственного ордена Дружбы народов медицинского университета В. Я. Бекиш; каф. биологии Гомельского государственного медицинского университета

Медицинская биология = Medical Biology : учебно-методическое пособие. М42 В 2 ч. Ч. 1 / В. В. Григорович, Ю. И. Корбут, Е. А. Черноус [и др.]. – 2-е изд., испр. – Минск : БГМУ, 2025. – 224 с.

ISBN 978-985-21-2059-3.

Содержит теоретический материал по программному материалу дисциплины «Медицинская биология и общая генетика» (разделы «Молекулярная биология», «Генетика»). Первое издание вышло в 2023 году.

Предназначено для иностранных студентов 1-го курса всех специальностей, обучающихся на английском языке.

УДК 57:61(075.8)

ББК 28.0я73

Учебное издание

Григорович Виктор Васильевич

Корбут Юлия Игоревна

Черноус Евгений Анатольевич и др.

МЕДИЦИНСКАЯ БИОЛОГИЯ

MEDICAL BIOLOGY

Учебно-методическое пособие

На английском языке

В двух частях

Часть 1

2-е издание, исправленное

Ответственный за выпуск В. В. Давыдов

Компьютерный набор Ю. И. Корбут

Компьютерная вёрстка Н. М. Федорцовой

Подписано в печать 30.10.25. Формат 60×84/16. Бумага писчая «Снегурочка».

Ризография. Гарнитура «Times». Усл. печ. л. 13,02. Уч.-изд. л. 13,62. Тираж 57 экз. Заказ 761.

Издатель и полиграфическое исполнение: учреждение образования

«Белорусский государственный медицинский университет».

Свидетельство о государственной регистрации издателя, изготовителя,

распространителя печатных изданий № 1/187 от 24.11.2023.

Ул. Ленинградская, 6, 220006, Минск.

ISBN 978-985-21-2059-3 (Ч. 1)

ISBN 978-985-21-2060-9

© УО «Белорусский государственный
медицинский университет», 2025

CHAPTER 1

MEDICAL BIOLOGY AND ITS ROLE IN MEDICAL EDUCATION. THE SUBJECT, TASKS, AND METHODS OF CYTOLOGY

THE NATURE OF LIFE, AND THE ROLE OF PROTEINS AND NUCLEIC ACIDS IN THE ORGANIZATION OF LIVING SYSTEMS

There are many different definitions of life. One of them was given by NASA: life is a self-sustaining chemical system capable of Darwinian evolution. However, it is possible to distinguish living from nonliving by a number of signs.

The fundamental properties of life that distinguish it from non-life are:

- Living organisms can create new individuals of their kind (self-reproduction).
- Living organisms can adjust themselves the environment (self-regulation).
- Living organisms can restore injured, lost, or worn-out parts (self-renewal).

The fundamental properties of living systems determine the main features of life:

- Life that we know is based on the functioning of *organic molecules* the most important of which are *nucleic acids* and *proteins*. Though, living organisms also have inorganic molecules and organic molecules can be produced in non-living nature.

- Living organisms *consist of cells*.

- Life is not possible without the *exchange of substances and energy* between the environment and an organism.

- Living organisms can multiply. The creation of new individuals of the same species by already existing organisms is called *reproduction*. “Multiplication” of living organisms is different from that of nonliving (e.g. fire) because of *heredity*.

- *Heredity* is the transmission of characteristics from parents to progeny. Because of heredity, children are similar to their parents.

- *Variation* is the ability of an organism to acquire new traits (ability to change). Due to this, children are never identical to their parents.

- Reproduction, heredity, and variation cause *evolution*. The evolutionary history of species development is its *phylogenesis*.

- Living organisms can *grow*. The development of an organism from the moment when its first cell is created to the death of the organism is called *ontogenesis*.

- *Irritability* is the capability of living matter to respond to environmental stimuli. For example, the nervous system causes the pupil to become smaller

when the eye is exposed to bright light. Such a response of organisms that have a nervous system is called a *reflex*. The movement of single cell to/from a stimulus (e.g. a higher concentration of a substance) is called *taxis*.

- *Homeostasis* is the dynamic constancy of an organism's internal environment (constant concentration of solutes in the blood, pH and etc.). *Integrity* means that the organism is functionally indivisible and should be considered as a solid functioning unit. *Discretion* means that any organism is composed of various parts and can be considered as a group of structural and functional units.

Understanding the chemical basis of life is necessary for understanding the mechanisms of genetic and other diseases.

The only life we know of is based on carbon biopolymers. The most important biopolymers are *RNA*, *proteins*, and *DNA*.

Proteins are the main “workers” of the cell. Virtually every action that takes place in the cell occurs with the participation of proteins. Proteins are long chains of amino acids (there are 20 types of amino acids with different properties (hydrophobic/hydrophilic, positively/negatively charged)). A single protein can consist of hundreds of amino acids, so their diversity is very high.

DNA is the molecule that stores information about the structure of all proteins in the cell. Proteins have a complex structure so they can be produced only with instruction. DNA consists of two long chains of four types of nucleotides. The order of nucleotides is the “code” that specifies proteins.

RNA is a molecule that has dual nature: it can catalyze reactions (like proteins) and store information (like DNA). It also consists of four types of nucleotides. The interactions of many proteins (and RNA) with various molecules and with each other direct all chemical and physical processes in the cell (and in its surroundings).

Organization levels of living matter:

1. Level of molecules. Processes occurring inside living organisms are chemical reactions — interactions and changes of different *molecules*. Molecules form *cells*.

2. Level of the cell. As mentioned above, all living organisms are composed of one or more cells. Cells are the smallest units of life.

3. Level of tissue. The organisms consisting of more than one cell, have tissues — groups of cells and extracellular matrix that are structurally and functionally related.

4. Level of organs. Tissues form organs — the body structures performing specific functions. Organs participating in the performance of the same major function are united into organ systems.

5. Level of the organism. All the organ systems form an organism. An organism is an elementary unit of life.

6. Level of population. A group of individuals of one species, occupying a definite territory for a long time, freely crossing and relatively isolated from other groups of individuals of the same species, form a population. A population is an elementary unit of evolution.

7. Level of biogeocenosis (community) and biosphere. A biogeocenosis (or community) includes all populations of different species which live on definite residential territory and are historically related to each other. All the communities compose the biosphere — an area of the planet occupied by living organisms.

HUMAN AS A BIOLOGICAL AND SOCIAL BEING

The human as a species refers to the phylum Chordates, subphylum Vertebrates, class Mammals, subclass Placentals, order Primates, suborder Anthropoids (narrow-nosed apes), family of Hominids (humans), genus Homo (man) and species Homo sapiens (wise man). Humans have characteristics of biological and social beings (Table 1).

Table 1

Similarity of humans and animals

Taxons of animals	Signs characteristics of human
Phylum chordates	Appearance of axial organs occurs in the embryonic period: a notochord, nerve tube, alimentary tube
Subphylum vertebrates	The notochord transforms into the spine, the heart is at the abdominal side. There are 2 pairs of extremities, 5 regions of the brain, and jaws
Class mammals	4-chambered heart, homoiothermy, well-developed cerebral cortex, mammary glands, presence of hair on skin coverings
Subclass placentals	Development of the human fetus in the mother's womb and its feeding through the placenta
Order primates	The thumb of the upper extremities is opposed to the others, nails on fingers, one pair of mammary glands, well-developed clavicles, teeth of three types and replacement of milk teeth by permanent ones, giving birth to one child in the majority of cases

The following signs are characteristic only of the species of Homo sapiens: straight walking, apparent thumb opposition, S-shaped spine, the brain volume of 1100–1700 cubic centimeters, prominence of the chin, abstract thinking, speech, production of instruments of labor and etc. The progress of humankind is controlled by social laws. Human life is impossible beyond society. Social factors have played a great role in human development. Knowledge, skills, and spiritual valuables are transferred to the society through training and education of young generations.

THE ROLE OF BIOLOGY IN MEDICAL EDUCATION

Biology has great significance for medical education. The important methods investigating the etiology (cause) and pathogenesis (development) of many human diseases and elaborating ways of their prevention and treatment are methods of biological modeling. Studying the biology of parasites is necessary for successful therapy (treatment) and prophylaxis (prevention) of invasive diseases. The World Health Organization suggests that about 90 % of all people are infected with helminthiases. Studying the mechanisms of cell physiology is crucial for understanding the action mechanism of many drugs and elaborating on new medicines. Genetic engineering (genetic designing of cells and organisms with definite characters) and biotechnologies (technological processes using living organisms) helped to set up the production of antibiotics, interferon, some hormones, enzymes, and vitamins. Gene therapy is a promising trend in the treatment of many hereditary disorders that could not be treated before. It is not possible without understanding the mechanisms of gene expression and the work of the human genome.

CYTOLOGY

Cytology is a science studying the structure, chemical composition, and functions of cells, their multiplication, development, and interaction in a multicellular organism. Tasks of cytology are:

- studying the structure and function of cells and their components (membranes, organelles, inclusions, and nuclei);
- studying the division of cells and possibilities of their adaptation to environmental changes;
- studying the interactions between cells in multicellular organisms.

The methods of cytology. Microscopy. Cells were discovered by Robert Hooke in 1665 with the microscope he constructed. Thus, microscopy was the first method of cytology. The reason humans are unable to see such small objects as cells is the limited *resolving power* of the human eye. Resolution is the ability of the optical system to distinguish two closely spaced points as separate objects. Another definition of resolution is the minimal distance at which two points are distinguishable by the optical system as two different objects. If two objects are closer than 100 μm apart, the light reflected from them reaches the same photoreceptor of the eye. If the objects are farther apart, the light from each object is detected by different cells, allowing the eye to perceive them separately. Microscopes help to overcome this limitation. Modern light microscopes can distinguish objects separated by at least 200 nanometers. However, they are not useful in studying smaller structures, such as the cell membrane. Adding more lenses to the microscope does not help to see smaller objects. The resolution of a light microscope cannot be less

than half the wavelength of visible light (approximately 0.4–0.7 μm). Under such conditions, light beams reflected from neighboring objects begin to overlap each other. To study smaller objects, electrons can be used instead of light. Electrons have much shorter wavelengths, so the electron microscope has a 1000 times greater resolution than a light microscope. Electron microscopes can distinguish objects up to 0.2 nm apart.

Light microscopy. An optical microscope is intended for studying micro-objects in a current of passing light. Light microscopes (Fig. 1) consist of mechanical, illuminating, and optical parts.

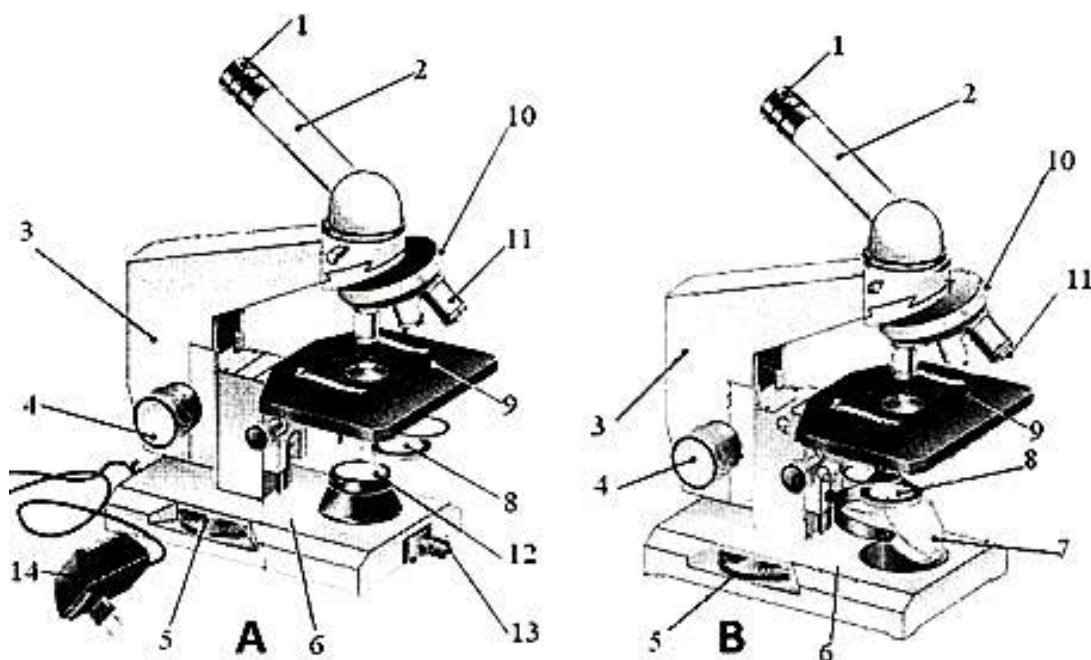


Figure 1. Light microscope:

1 — ocular lens; 2 — drawtube; 3 — arm (drawtube holder); 4 — coarse adjustment knob (cremaliera); 5 — fine adjustment knob; 6 — base; 7 — mirror; 8 — condenser; 9 — stage; 10 — revolving nosepiece; 11 — objective lenses; 12 — lens collecting light; 13 — lamp; 14 — adapter

The mechanical part includes the base, stage, coarse adjustment knob (cremaliera), fine adjustment knob, arm, and revolving nosepiece. The support of the microscope consists of an arm (draw-tube holder) and a base. The arm carries:

- revolving nosepiece — a rotating mechanism for changing objective lenses;
- draw-tube — a hollow tube with an ocular lens;
- a system of knobs for coarse and y adjustment of the microscope;
- a stage for placing an investigation object.

The illuminating part allows us to illuminate the object and make it visible. It includes a lamp (or mirror in simpler microscopes) and a condenser. The mirror of the microscope is double-sided and has flat and concave surfaces.

The concave surface is used under natural illumination, while a flat one is under intensive artificial illumination. The condenser is a lens system collecting light rays into a band.

The diameter of the light beam can be regulated with a small lever which changes the diaphragm lumen.

The optical system consists of one ocular lens (or two in binocular microscopes) and several objective lenses with different magnifications.

The ocular lens (oculus — eye) is a lens system directed toward the eye. Its magnification is indicated on the frame. A training microscope has spare ocular lenses with magnifications of $7\times$, $10\times$, and $15\times$.

The objective lens is located on the nosepiece at the lower end of the drawtube. It is a lens system directed to the investigated object. Two kinds of objective lenses are commonly used in a training microscope: with low ($8\times$) and high magnifications ($40\times$).

The total magnification of the microscope can be calculated by multiplying the magnifications of the objective and ocular lenses. For example, if the magnification of an objective lens is $40\times$ and that of an ocular lens is $7\times$, then the total magnification of the microscope is equal to $280\times$.

HISTOCHEMISTRY AND IMMUNOHISTOCHEMISTRY

Fluorescence microscopy. Common dyes are distributed throughout the cell and stain many structures nonspecifically, but there are special dyes that stain only certain substances. Such differential staining of cells is called *histochemistry*. Histochemical reactions reveal the sites of specific chemical reactions or specific cellular components.

Examples of histochemical reactions are the Feulgen reaction, which stains DNA red, proteins with bromophenol blue, and lipids with sudan black (Fig. 2).

Fluorescent dyes (e.g., acridine orange, which stains nucleic acids) are also used in histochemistry (Fig. 3).

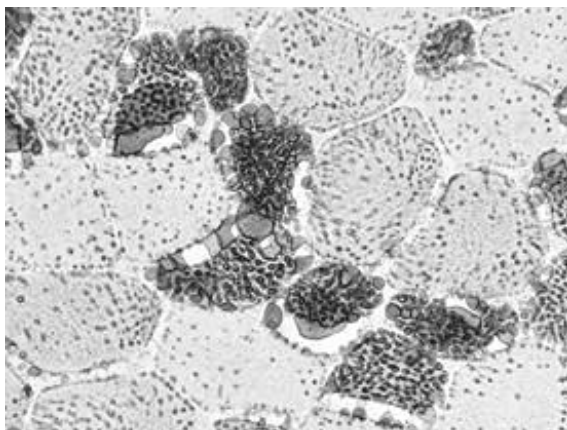


Figure 2. Staining of lipids with sudan black

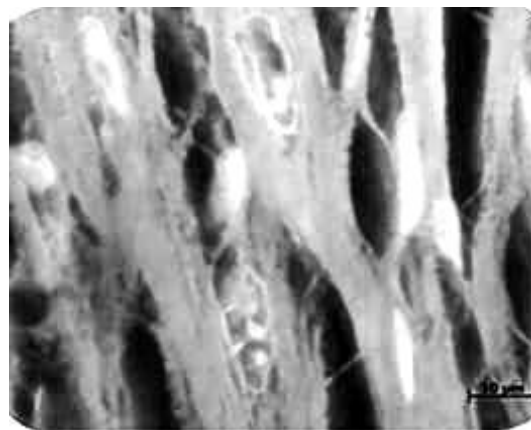


Figure 3. Nuclei stained with acridine orange

Fluorescent dyes are dyes that absorb light of one wavelength and re-emit light at a longer wavelength. Fluorescence microscopy (also called luminescence microscopy) is used to obtain images from such dyes. A fluorescence microscope is a type of light microscope. The studied sample is with a lamp, the light passes through a filter which leaves only the wavelengths that are optimal to cause fluorescence of the dye. Thus, only the light from the fluorescent dye, not from the bright lamp can be observed by microscopists. *Immunohistochemistry* allows detection of specific proteins with antibodies bound with dyes (Fig. 4). Antibodies are proteins of the immune system that can recognize specific antigens (usually proteins) and bind to them.

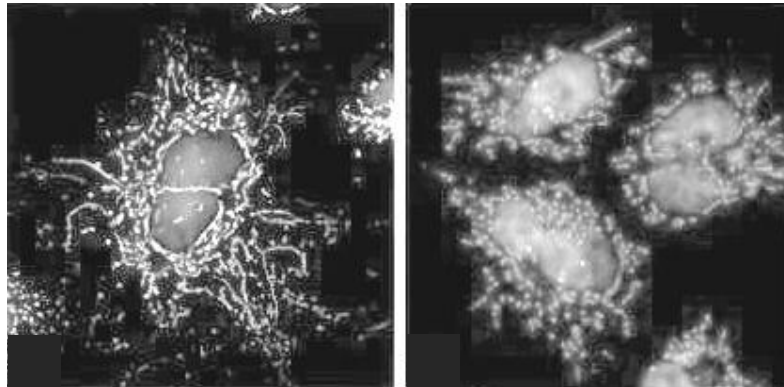


Figure 4. Mitochondria stained with monoclonal antibodies

Morphometry. Morphometry is a set of methods used to measure the size or count the number of histological structures. Measurements can be made with the help of ocular micrometers. Ocular micrometers are glass discs with a scale or measuring grid on them (Fig. 5). They are placed into the microscope. Modern devices allow measurements in automatic and semi-automatic modes.

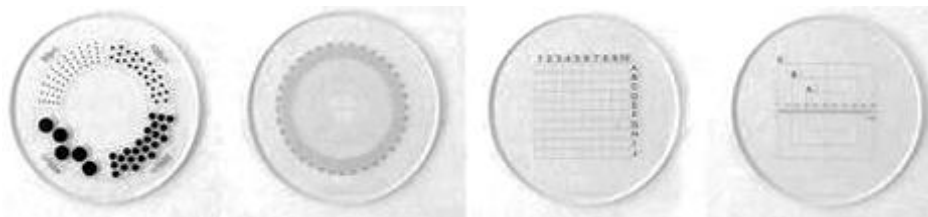


Figure 5. Ocular micrometers

Electron microscopy. The structure of an electron microscope is similar to that of an optical one. Electron microscopy uses electrons instead of photons of visible light. The resulting images are black and white, but they are often colored to show specific structures.

Transmission electron microscopy (TEM) is based on passing a beam of electrons through an ultra-thin slice of the sample under study (Fig. 6). Some of the electrons are absorbed by the sample or reflected. Others pass through the sample and fall onto a screen.

Scanning electron microscopy (SEM). A beam of electrons scans the surface of the sample. The electrons reflected from the surface are detected to produce the image of the sample. Scanning electron microscopy produces high-quality three-dimensional images of the surface of objects (Fig. 7).

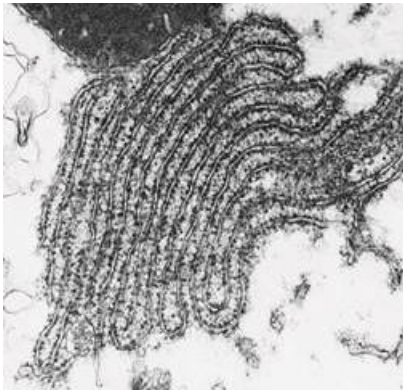


Figure 6. TEM of endoplasmic reticulum

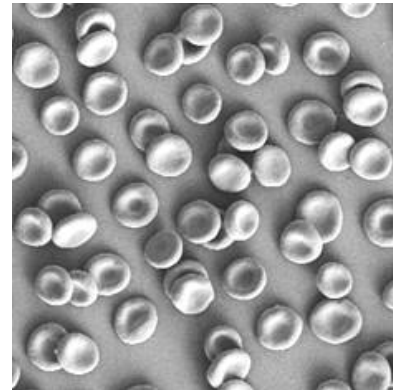


Figure 7. SEM of red blood cells

Cryo-electron microscopy (cryo-EM). Tissue samples or aqueous protein solutions are rapidly frozen at temperatures below -160°C , fixing the molecules in their natural state. Then electron microscopy is used to visualize the molecules and create a three-dimensional image. Recent advances in technology have made it possible to determine the structures of biomolecules with almost atomic resolution (Fig. 8). In 2017, the Nobel Prize in Chemistry was awarded for the development of this method.

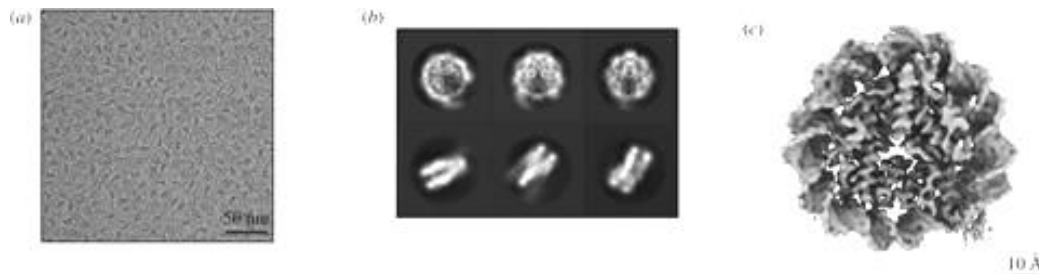


Figure 8. Cryo-EM of a nucleosome

Cell fractionation (differential centrifugation). This method is used to separate (fractionate) cell components according to their size and density. The cells are crushed in a homogenizer to break them up. The resulting mixture is centrifuged. This causes the heavier components to precipitate to the bottom of the test tube. The liquid over the precipitate is poured into another tube and centrifuged at a higher speed for a longer time. This process is repeated several times. These procedures produce a set of precipitated fractions, each containing different cellular components (Fig. 9).

This method is used for extracting different cellular components such as nuclei, ribosomes and many others. An advanced method, ultracentrifugation, can even separate different protein fractions.

Isotopic labeling. This method detects radiation from isotopes contained in an object of study with a photosensitive film. It helps to study the tissue distribution of a molecule in the sample. Examples of radioactive tags are ^{131}I , ^{35}S , ^{14}C , and ^3H .

Cell Cultures. A cell culture is cells of any tissue grown in a laboratory dish in a special nutrient medium under controlled conditions.

Applications of cell cultures:

- General biological research. Cell cultures are used as models to study cell biology, the interaction between cells and pathogens such as bacteria and viruses, the effects of drugs, and the aging process.

- Cancer research. The basic differences between normal and cancer cells can be studied using animal cell cultures. This attitude helps to study the mechanisms of carcinogenesis, and tumor development and develop effective anti-cancer drugs.

- Drug development and testing. Animal cell cultures are used to study the cytotoxicity of new drugs and to determine their effective and safe doses.

- Virology. Cell culture can also be used to study the life cycle of viruses.
- Vaccine production. Cultures of animal cells can be used to produce viruses for manufacturing vaccines. For example, vaccines for polio and hepatitis B can be produced with animal cell culture. Vector vaccines against COVID-19 are also produced on cell cultures.

- Genetic engineering. Animal cell cultures are used to produce many important genetically engineered proteins, such as monoclonal antibodies, insulin, and hormones (see chapter “Genetic engineering”).

- Gene therapy. Culturing the patient’s cells is a step of ex vivo gene therapy (see chapter “Molecular genetics in Medicine”), where the patient’s cells are obtained, and the genetic defect in them is corrected and reintroduced to the patient.

- Genetic counseling. Fetal cell culture obtained from a pregnant woman can be used to diagnose genetic diseases and chromosomal abnormalities in the fetus (see chapter “Genetic counseling”).

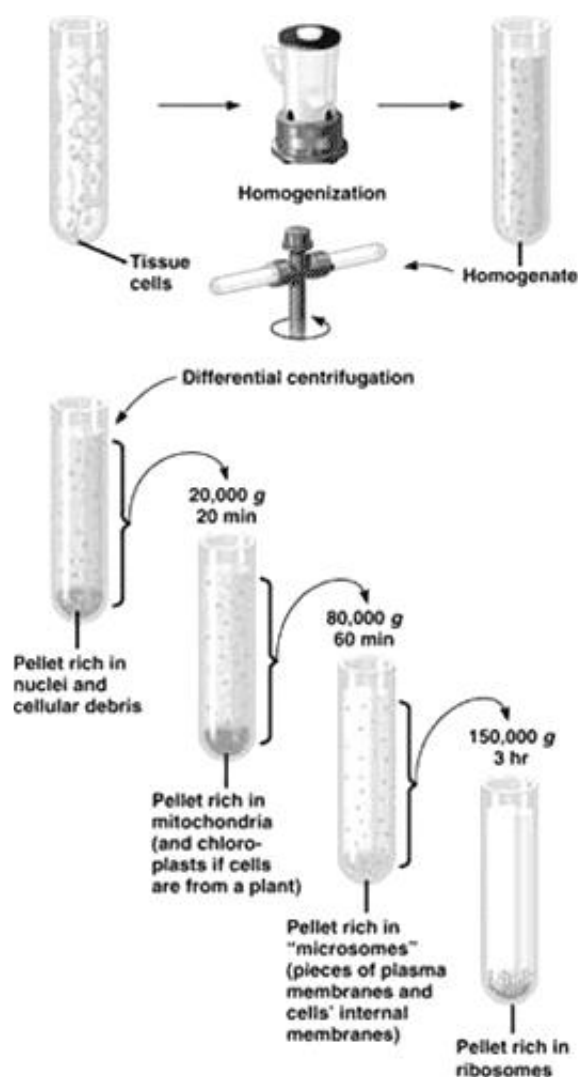


Figure 9. Differential centrifugation

- Transplantology. Many research groups work on the creation of artificial organs such as the liver, kidney, and pancreas from stem cells.

Cell cultures can be produced from any tissue of the body:

- The specimen of tissue is obtained.
- The extracellular matrix (a complex of protein filaments and polysaccharides in the intercellular medium to which the cells are attached) is dissolved and the bonds between the cells are broken.
- The cells are then placed in a dish with a nutrient medium.

Most cells can divide a limited number of times (40–60). The number of times that a normal population of somatic differentiated human cells can divide is called the Hayflick limit. However, this limit does not exist in cancer cells and they are often used in laboratories. An example of such a culture is HeLa cells. They were obtained in 1951 from Henrietta Lacks, who died of cervical cancer the same year.

Other methods. Many other methods are used in cytology, among them:

- *Biochemical methods* are used to study the chemical composition of cells.
- *X-ray crystallography* allows determining the three-dimensional structure of biomolecules. This method is based on the diffraction of X-rays passing through the crystal of the substance under study. Mathematical processing of the obtained data in combination with knowledge about the chemical properties of the substance helps to determine its three-dimensional structure.

- *Microsurgery* of cells is the method allowing removing and transplanting of organelles of the cell

THE CELL THEORY

The components of cells are:

1. Plasma membrane is the covering of the cell consisting of lipids (they will be considered in the next chapter).

2. Nucleus is the cell's compartment which contains the information that helps to produce proteins — the molecules very necessary for life. Proteins will be considered in the next chapter. The nucleus is present in cells of animals, plants, fungi, and protists (these organisms are called *eukaryotes*), but not bacteria (they belong to *prokaryotes*).

3. Cytoplasm — all the contents of the cell between the membrane and the nucleus. Cytoplasm contains *organelles* — differentiated structures performing particular functions. The rest of the cytoplasm is cytosol — viscous water solution of organic and inorganic molecules.

All basic knowledge about cells was expressed in the Cell Theory in 1839:

1. Cells are the smallest living units of all living organisms.
2. Cells of animals, plants, and other organisms are similar in structure, chemical composition, and functioning.

3. Growth and development of organisms occur due to the multiplication and development of cells.

4. New cells appear only by means of the division of already existing cells.

Prokaryotes and eukaryotes. Based on the structural organization, there are two major kinds of cells: prokaryotes and eukaryotes.

Prokaryotes include *archaea* and *bacteria*. A prokaryotic cell is a cell that has no membrane-bound organelles (i.e., covered by a membrane) and no nucleus. Their DNA is in the cytoplasm. Prokaryotes are not capable of phagocytosis, they are usually unicellular, commonly much smaller than eukaryotic cells. Prokaryotes are usually haploid — they have only one chromosome. Due to the simpler organization of the genome, they do not have mitosis and multiply by binary fission. Prokaryotes do not have sexual reproduction and meiosis, usually have cell walls.

Eukaryotes include protists, fungi, plants, and animals. Key features of a eukaryotic cell are the presence of a membrane-enclosed nucleus and membrane-bound organelles. The DNA is not a single molecule, but several or many molecules contained in chromosomes. Eukaryotes can be either haploid or diploid. The structure of the cell wall differs from that of prokaryotes. Animals do not have a cell wall.

Viruses. A virus is an infectious agent consisting of nucleic acid (DNA or RNA) enveloped into proteins and in some cases lipids (Fig. 10). The outer protein shell of viruses is called a capsid. An entire virus particle, consisting of a capsid and an inner core of nucleic acid is called a *virion*. Viruses bring their genome into the cell to use for reproduction. In other words, they are obligate intracellular parasites. Viruses are not cells, they do not have the characteristics of the living matter until they enter the cell, cannot be grown on nutrient media, cannot divide (they self-assemble from components synthesized in the cell), do not have ribosomes. There are viruses of eukaryotes, bacteria, and archaea.

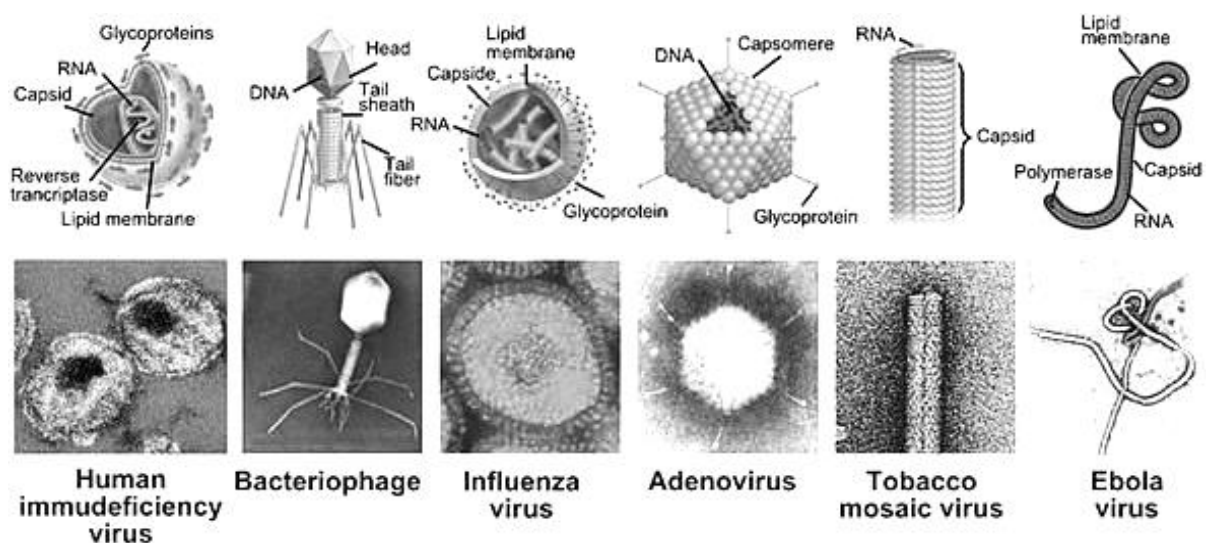


Figure 10. Viruses

The genomes of different viruses may contain RNA or DNA, which can be single-stranded or double-stranded. The DNA or RNA can be linear or circular (except for double-stranded RNA which is always linear). The genome can be segmented (represented by more than one molecule). The number of genes in a genome is usually several to several hundred (in eukaryotes, thousands).

The protein envelope surrounding the nucleic acid of a virus is called a capsid. The capsid protects the viral genome from external influences and takes part in viral entry into the cell. It consists of multiple repeating proteins called capsomers. The capsomers fit each other to assemble a capsid. The capsid can make a virus rod-shaped or nearly spherical (rarely other shapes are possible). Some viruses (herpesviruses, influenza virus) are also covered with a lipid envelope called. It originates from the membrane of the infected cell (or the membrane of the nucleus). The lipid envelope also contains viral glycoproteins. Typical virus life cycle: attachment of the viral particle to the cell surface → entry of the virus into the cell → synthesis of viral proteins nucleic acids → viral particles leave the cell.

There are three possible patterns of viral genome replication:

1. DNA → DNA. If the DNA is double-stranded, its replication is similar to that of the host cell's DNA. They commonly use DNA polymerase of the host cell.

2. RNA → RNA. Infected cells do not have the enzyme that can replicate RNA and most RNA viruses contain a gene encoding RNA replicase.

3. RNA → DNA → RNA. Some RNA viruses encode reverse transcriptase, an enzyme that makes DNA on the RNA template.

The viral nucleic acid and capsid proteins spontaneously assemble into new viral particles, a process called self-assembly. For example, the tobacco mosaic virus can be “disassembled” in the laboratory. When the components of the virus are mixed, they spontaneously assemble into virions.

Transposons. Transposons are DNA sequences that can move from one location of the genome to another. They are found in both prokaryotes and eukaryotes. They are of two types:

- Transposons (or DNA transposons), which can move by a “cut-and-paste” mechanism.

- Retrotransposons, which can move by the “copy-paste” mechanism with reverse transcription.

The enzymes necessary for this are encoded by the transposon itself.

Viroids and prions. Another class of plant pathogens is viroids. They are small, circular RNA molecules that do not encode proteins but replicate in the host plant cells. Viroids probably disrupt normal plant metabolism, by causing errors in the systems that control gene expression.

Prions are defective (misfolded) forms of normal cellular proteins. When prions infect normal cells, they convert normal proteins into prions.

CHAPTER 2

STRUCTURAL AND FUNCTIONAL ORGANIZATION OF THE CELL

THE STRUCTURE OF THE PLASMA MEMBRANE

In 1935, H. Dawson and J. Danielli proposed the sandwich model of the membrane: the lipid bilayer is enclosed between two layers of proteins. Such a model agreed with some data, such as the surface tension of the lipid bilayer. However, with time, data contradicting this model were collected.

In 1972, S. Singer and G. Nicholson proposed the fluid mosaic model of the membrane: proteins do not cover the membrane, but are individually embedded in the lipid bilayer (Fig. 11). The hydrophilic parts of the proteins are maximally exposed to water, while the hydrophobic parts of the proteins are in a non-aquatic environment inside the lipid bilayer. Thus, proteins drift in a “lipid sea.” The understanding of the mechanism forming plasma membrane is based on the following facts:

- Water molecules, the main component of cell and extracellular space, are polar and therefore easily interact with other polar or charged molecules. Hence, water can dissolve polar or charged molecules but does not dissolve non-polar molecules.
- Because of this, non-polar molecules tend to stay close to each other, reducing the contact area with water.
- Phospholipid molecules, the most common lipids of any cell membrane, consist of a polar (also called hydrophilic) head and nonpolar (i.e. hydrophobic) tails.
- In aquatic solution, phospholipids form double layers, closed into bubbles. This position of phospholipids “hides” their nonpolar tails from the water, while allowing their polar heads to contact the water surrounding the membrane.

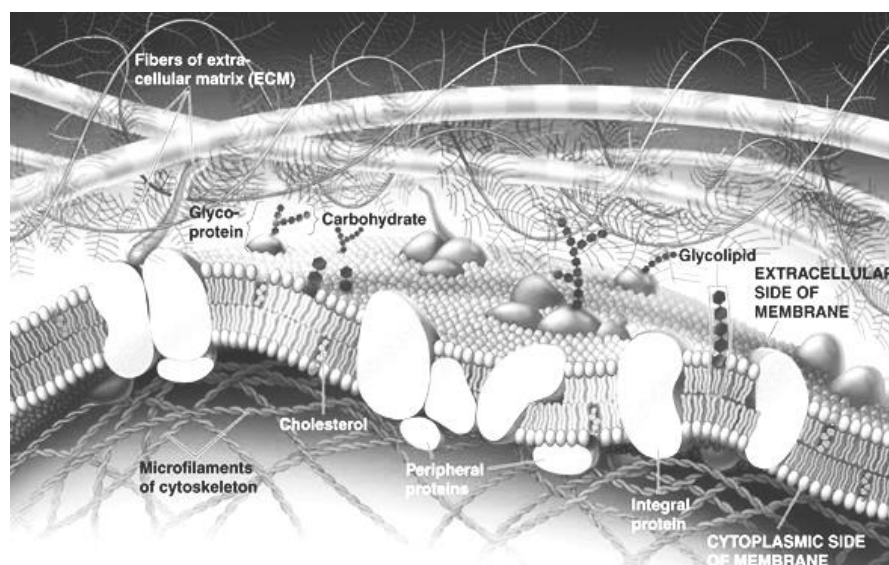


Figure 11. The structure of the membrane of an animal cell

These features allow phospholipids to be held together without covalent bonds and provide the following properties of the plasma membrane: plasticity, self-locking, and selective permeability.

In addition to phospholipids, the plasma membrane contains sphingolipids, cholesterol, and some other lipids, as well as proteins. Membrane proteins differ both in structure and function. According to their spatial arrangement, they can be divided into two main groups:

1. Integral proteins are proteins with hydrophobic sites completely enclosed in the hydrophobic part of the plasma membrane. Integral bitopic and polytopic proteins penetrate the phospholipid bilayer once or several times, respectively. Integral monotopic proteins (also known as semi-integral proteins) are immersed in only one lipid layer.

2. Peripheral proteins are not embedded in the lipid bilayer but attach to its surface.

Membrane proteins can perform various functions, the most typical of which are:

- Transport proteins pump molecules across the membrane.
- Receptor proteins interact with specific signaling molecules and cause a cell response.
- Membrane enzymes catalyze specific chemical reactions.
- Some proteins bind to the extracellular matrix or other cells.

There is also a protein network beneath the eukaryotic cell membrane. It reinforces and supports the membrane.

The membrane is also bound to carbohydrates. Membrane carbohydrates of the animal cell are usually short branched chains consisting of less than 15 monomers. Carbohydrates bound to lipids are called glycolipids, and carbohydrates bound to proteins are called glycoproteins. Glycolipids and glycoproteins form a jelly-like layer on the outer surface of the membrane. This layer functions as a barrier between the cell and its environment, protecting the cell membrane from the direct action of physical forces. It also participates in intercellular interactions. Carbohydrate chains may function as markers that distinguish one type of cell from another. For example, ABO blood groups (groups I–IV) reflect differences in the carbohydrate of certain glycoproteins on the surface of red blood cells.

Bacteria have a cell wall consisting of murein (peptidoglycan) — a network of polysaccharides linked by short peptide chains. It gives the cell its shape and protects it from rupture in a hypotonic solution. The mechanism of action of many antibiotics, such as penicillin, is related to the disturbance of peptidoglycan synthesis. Since peptidoglycan is characteristic of bacteria, such antibiotics do not affect human cells.

TRANSPORT ACROSS THE PLASMA MEMBRANE

The concentration of many substances is not the same on either side of the membrane. The value describing the difference in concentration of a dissolved substance in a solution and the direction of its change is called a *concentration gradient*. The transport of substances across the membrane can occur both down the concentration gradient (i.e. from the area of higher solute concentration to the area of its lower concentration) and against it (from the area of lower solute concentration to the area of higher concentration). In this regard, two main types of transport are distinguished: passive and active.

Passive transport. Passive transport: molecules of the dissolved substance move down the concentration gradient; no energy input is required; the concentration difference of the dissolved substance gradually decreases.

Passive transport includes simple diffusion, facilitated diffusion, and osmosis (Fig. 12). Diffusion is the passive transport (without energy expenditure) of dissolved substances across the membrane down the concentration gradient. Diffusion rate increases with increasing temperature, membrane surface available for transport, and concentration difference.

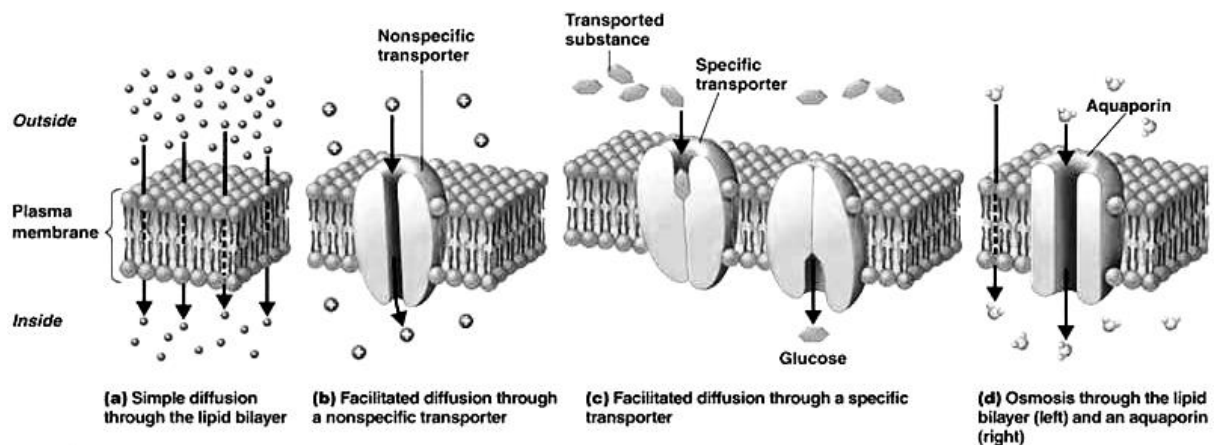


Figure 12. Passive transport

Simple diffusion — molecules are transported through the membrane passing between phospholipids. Substances soluble in lipids, such as steroid hormones, or small nonpolar molecules such as O_2 and CO_2 are transported this way. Small polar uncharged molecules such as ethanol can also pass through the membrane similarly.

Facilitated diffusion is the diffusion of dissolved substances through transport proteins. Ions (e.g. Na^+ , H^+) as well as larger polar molecules (e.g. glucose) can pass through the membrane in this way.

Osmosis is the diffusion of water through a semipermeable membrane into an area with a higher concentration of solutes. If there are two solutions with different solute concentrations, the solution having a higher concentration of solutes is called a hypertonic solution. The solution having a lower

concentration of solutes is called a hypotonic solution. The solution with the same solute concentration as the other one is called an isotonic solution.

When placed in a hypertonic solution, cells lose water by osmosis. In a hypotonic solution, water enters the cell. In an isotonic solution, the inflow of water into and out of the cell is equal. Therefore, isotonic solutions (0.9 % sodium chloride solution and 5 % glucose solution) are used for intravenous injection.

During osmosis, water can enter the cell through the lipid bilayer, but the rate of such current is limited. The cell has protein channels for water transport. They are called aquaporins.

Active transport. Molecules of the dissolved substance move from the area of lower concentration to the area of higher concentration (i.e., against the gradient). Energy expenditure is required. Transport occurs with the participation of special proteins. The concentration difference increases.

The proteins performing active transport bind only to certain types of molecules, such as carbohydrates, amino acids, nucleotides, and ions. One protein may carry more than one type of molecule (Fig. 13):

- Uniport is the transport of one type of molecule in a particular direction.
- Symport is the transport of two different kinds of molecules or ions together in the same direction by a transport protein.
- Antiport is the transport of two different kinds of molecules or ions in opposite directions by the same transport protein.

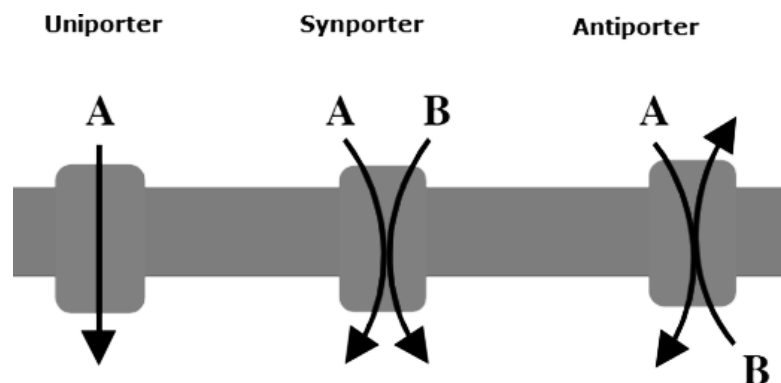


Figure 13. Active transport

An example of antiport is the Na^+/K^+ pump, which is present in all animal cells. This protein captures three Na^+ ions from the cytoplasm, hydrolyzes ATP to ADP, transfers them to the outside of the cell, and then transfers two K^+ ions in the opposite direction. This mechanism helps to maintain a high concentration of Na^+ outside the cell and a high concentration of K^+ in the cytoplasm.

The energy source for active transport can be either ATP hydrolysis (primary active transport) or the electrochemical gradient created by the difference in concentrations of a certain ion on either side of the membrane (secondary active transport). In this case, the necessary substance can be

transported against the concentration gradient together with the ion transported down the concentration gradient. For example, a specific protein in renal proximal tubule cells simultaneously transports one molecule of glucose together with two Na^+ ions into the cell.

Bulk transport (exocytosis and endocytosis). During exocytosis and endocytosis, transported substances in a relatively large amount are enclosed in membrane vesicles and transported outside or inside the cell. These processes are also energy-requiring and occur only in eukaryotes.

Exocytosis is the export of macromolecules enclosed in a membrane-bound vesicle from the cell by the fusion of this vesicle with the plasma membrane. For example, pancreatic cells secrete insulin and nerve cells release neurotransmitters by exocytosis.

Endocytosis is the process of importing macromolecules into a cell by coating them into the plasma membrane and forming vesicles (Fig. 14). The types of endocytosis are phagocytosis, pinocytosis, and receptor-mediated endocytosis.

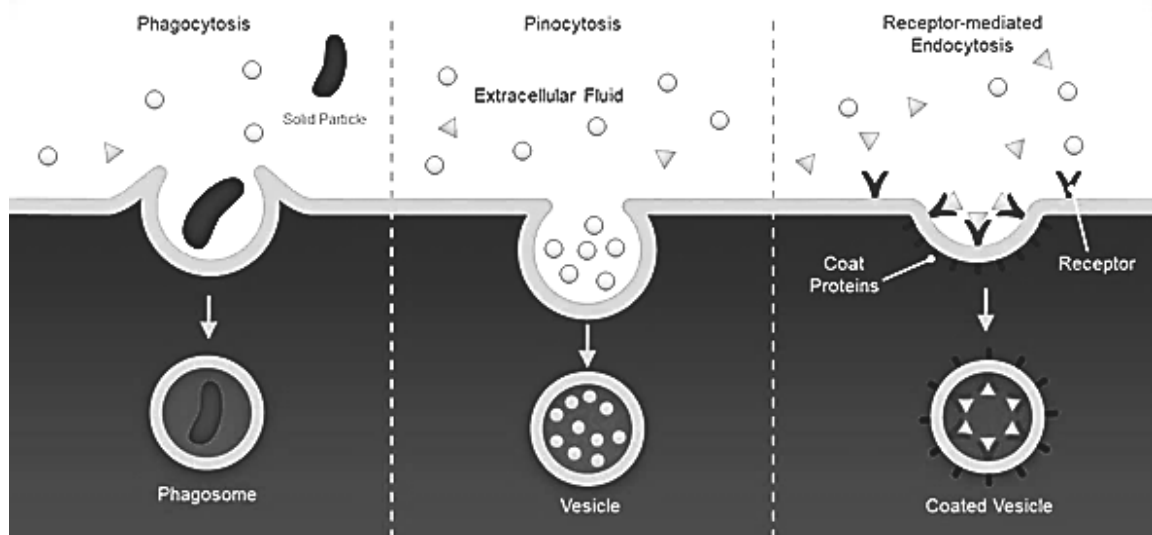


Figure 14. Endocytosis

During *phagocytosis*, the cell engulfs a solid particle surrounding it with pseudopodia and encloses it in a vesicle. Such a vesicle entering the cytoplasm is called a phagosome. An example of phagocytosis is the uptake of bacteria by cells of the immune system, such as neutrophils or macrophages.

Pinocytosis is the form of endocytosis in which the cell engulfs fluids. Pinocytosis is nonspecific for transported substances — all dissolved substances enter the cell.

In *receptor-mediated endocytosis*, a receptor on the cell surface binds to a specific substance. As a result, the membrane region where this has occurred forms a bud and then detaches, becoming a transport vesicle.

An example of such transport is the entry of cholesterol into cells, and defects or lack of necessary receptors due to a mutation causes a hereditary disease called familial hypercholesterolemia.

CYTOPLASM

The cytoplasm contains cytosol, organelles, and inclusions. Cytosol, or hyaloplasm, is the liquid part of the cytoplasm. The cytosol is a complex mixture of substances dissolved in water. Inclusions are formations that do not have an active function and are located in the cytoplasm. They may appear and disappear at different stages of the cell's functioning. Organelles are relatively isolated structures in the cell cytoplasm that perform specific functions.

Organelles. Organelles can be classified according to several criteria:

- Non-membrane bound organelles are not covered by the plasma membrane (ribosomes, centrosomes, myofibrils). Membrane-bound organelles (endoplasmic network, Golgi complex, mitochondria, lysosomes, etc.) are covered with a lipid bilayer with proteins.

- Organelles of general purpose are necessary for almost all cell types (endoplasmic reticulum, Golgi complex, lysosomes, etc.), organelles of special purpose are present only in some specialized cells (flagellum and acrosome of spermatozoa, cilia in respiratory cells).

- Organelles of the anabolic cell system (ribosomes, endoplasmic reticulum, Golgi complex) are mainly responsible for anabolic reactions, while organelles of the catabolic cell system (mitochondria, lysosomes, peroxisomes) — for dissimulation.

Ribosomes.

- Have no membrane.
- Synthesize proteins.
- Present in large amounts in both eukaryotic and prokaryotic cells.
- The ribosome of any organism consists of large and small subunits. Both of them are composed of proteins and rRNA and exist separately until they start protein synthesis. Ribosomes of prokaryotes contain 3 rRNA molecules and their sedimentation coefficient is 70s. Eukaryotic ribosomes have 4 rRNA, 80s. The number of proteins in the ribosome varies. For example, the human ribosome contains 80 proteins, 47 of which are in the large subunit and 33 in the small subunit. Ribosomes of mitochondria and plastids are similar to those of bacteria.

- In a eukaryotic cell, ribosomes can be located in the cytosol (free ribosomes) or attached to the outside of the endoplasmic reticulum (Fig. 15) or the outer nuclear membrane (bound ribosomes). Free ribosomes synthesize proteins that function in the cytosol, nucleus, and mitochondria. Bound ribosomes (Fig. 16) synthesize proteins of the cell's membrane, endoplasmic reticulum, Golgi body, lysosomes, peroxisomes, and proteins destined for

release from the cell (such as insulin synthesized in pancreatic cells). Any ribosome can become either free or bound.

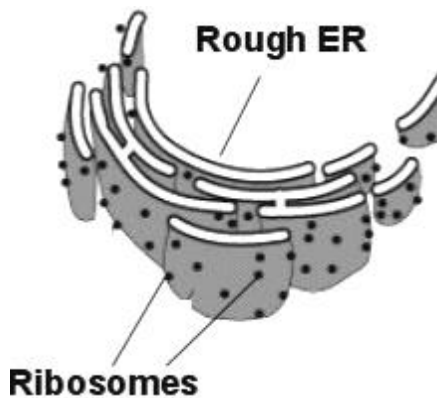


Figure 15. Ribosomes on ER

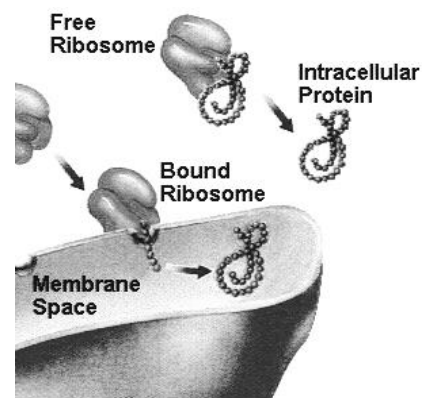


Figure 16. A ribosome binding to ER during protein synthesis

THE ENDOMEMBRANE SYSTEM

The inner part of a eukaryotic cell is filled with membranes forming an endomembrane system. It divides the cell into compartments. The presence of these membranes in eukaryotic cells is one of the fundamental differences between eukaryotes and prokaryotes. The endomembrane system includes the nuclear envelope, endoplasmic reticulum, Golgi apparatus, lysosomes, various types of vesicles and vacuoles, and the plasma membrane. Communication between these structures is provided either by the continuity of the common membrane that forms them (nucleus and endoplasmic reticulum) or by vesicle transport (endoplasmic reticulum and Golgi apparatus).

The endoplasmic reticulum (ER) is a large network of flattened cavities (cisternae) and channels separated from the cytosol with the membrane (Fig. 17).

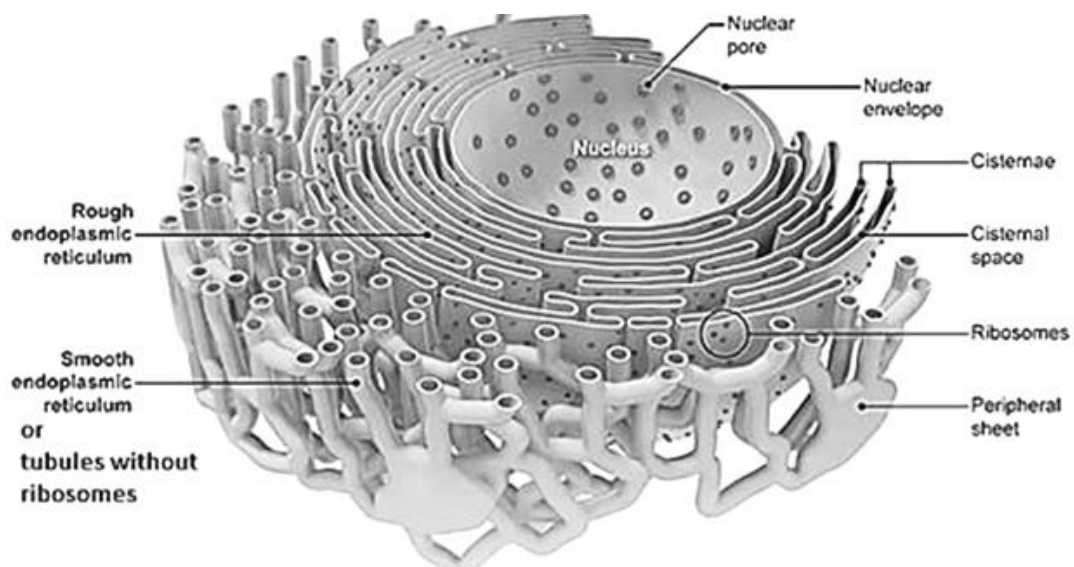


Figure 17. Diagram of ER

ER is usually concentrated mainly around the nucleus and occupies a considerable volume of cell cytoplasm. The ER membrane is an extension of the outer nuclear membrane.

Rough (or granular) ER consists mainly of flat cisternae, the surface of which looks rough due to the presence of ribosomes on it.

Protein synthesized by bound ribosomes enters rough ER through a special pore, where its primary modification and proper folding occur. Proteins that promote the proper folding of other proteins are called chaperones.

The processed proteins are transported to the Golgi apparatus in vesicles. Smooth ER is mainly formed by branching and merging tubes and channels. Membranes of rough and smooth ER are connected.

Smooth ER contains a large number of membrane-bound enzymes that synthesize various carbohydrates and lipids.

ER may serve as the storage of intracellular Ca^{2+} . For example, in muscle cells, the release of Ca^{2+} from the ER (called the sarcoplasmic reticulum in muscle cells) is used to trigger muscle contraction.

Smooth ER (especially in liver cells) recycles poisons and medicines by adding hydroxyl and other groups. Poisons and medicines become water-soluble and can be removed from the body with urine.

Golgi apparatus consists of many flat disc-shaped sacs called cisternae. Cisternae are thin in the center and considerably thicker at the periphery. The stack of cisternae of the Golgi apparatus is called dictyosome.

The Golgi apparatus participates in chemical rearrangements (attachment and modification of oligosaccharides), sorting, and accumulation of products synthesized in the ER. The two sides of a Golgi stack are referred to as the cis face and the trans face.

The cis face is usually located near the ER and accepts transport vesicles coming from it. The trans face produces new vesicles that transport products from the Golgi apparatus to their further destination.

The Golgi apparatus is also a source of lysosomes.

Lysosomes are membrane vesicles that contain hydrolytic enzymes which digest (hydrolyze) macromolecules (Fig. 18).

Lysosome enzymes are most active in an acidic environment, so they are not too dangerous if released into the cytosol, which has an almost neutral pH.

The components of lysosomes are produced in the ER and enter the Golgi apparatus.

Newly produced lysosomes (primary lysosomes) fuse with membrane vesicles that contain the substrate for digestion and form secondary lysosomes.

The process in which lysosomes fuse with a vacuole from the outside (phagosome) and digest its contents is called heterophagy. For example, the cells of the immune system ingest pathogenic microorganisms and process them with the help of lysosomes.

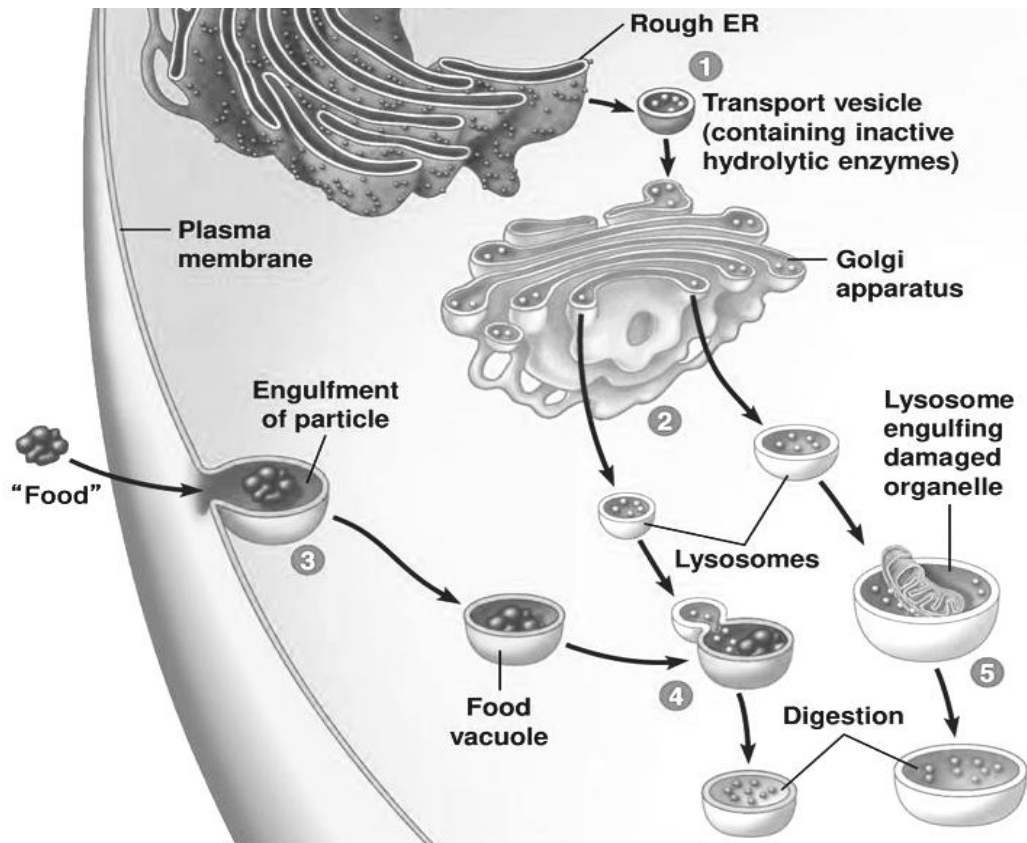


Figure 18. Biogenesis of lysosomes, autophagy and heterophagy

In the case of autophagy, the damaged organelle or some amount of cytosol is surrounded by a double membrane (its origin is unclear). The lysosome fuses with the outer membrane of the formed vesicle. The products resulting from digestion are transported to the cytosol through the lysosome membrane with the help of special receptors.

There are hereditary diseases that are characterized by a deficiency of lysosomal enzymes. It's resulting in the accumulation of unsplit macromolecules in the lysosomes. Such diseases are called lysosomal storage diseases.

Peroxisomes are vesicles 0.3–1.5 μm in diameter with the crystallized nucleus.

New peroxisomes are formed by the endoplasmic reticulum, but can also be formed by the division of existing peroxisomes.

Peroxisomes contain enzymes that use molecular oxygen to remove hydrogen atoms from organic molecules. This results in the formation of H_2O_2 .

Peroxisomes also contain the enzyme catalase, which breaks H_2O_2 .

An important function of peroxisomes is the oxidation of fatty acids to form acetyl-CoA, which is then exported to the cytosol.

Mitochondria are elongated organelles about 1–10 μm long (Fig. 19).

In living cells, mitochondria form a dynamic branching tubular network. They are constantly disintegrating into separate fragments, which are reassembled.

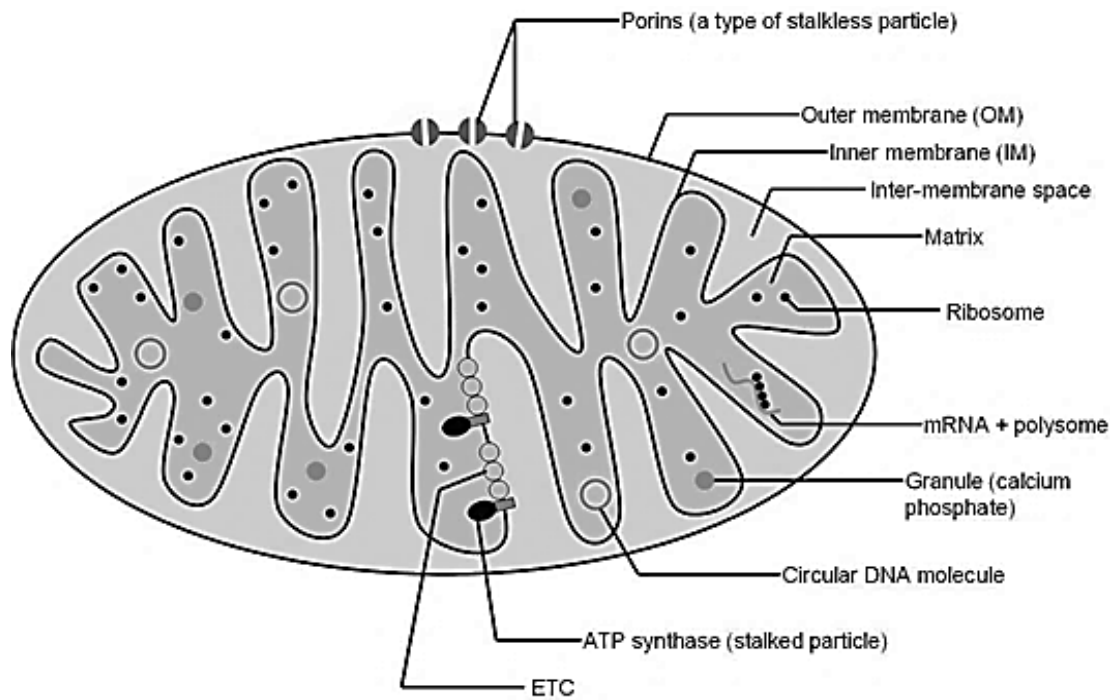


Figure 19. Diagram of a mitochondrion

Mitochondria have outer and inner membranes, the space between them is called perimitochondrial (intermembrane) space. The content of these organelles surrounded by the inner membrane is called the matrix.

In addition to various enzymes, the matrix contains ribosomes and circular mitochondrial DNA (mtDNA).

Human mtDNA consists of 16,569 base pairs (almost 200,000 fewer than the nuclear genome) and contains 37 genes encoding 13 proteins, 22 tRNAs, and 2 rRNAs.

Significant morphological and molecular-genetic similarities between mitochondria and bacteria suggest the origin of these organelles: “domestication” of bacteria by a protoeukaryotic cell.

The main function of mitochondria in the cell is ATP synthesis during which organic substances are broken down to CO_2 and H_2O .

CYTOSKELETON

The cytoplasm of all eukaryotic cells is crossed by a network of protein fibers called the *cytoskeleton*. The cytoskeleton is a constantly assembling and disassembling dynamic system. Individual filaments of the cytoskeleton are composed of polymers of repeating protein subunits. There are three main types of cytoskeleton filaments: actin microfilaments, intermediate filaments, and microtubules (Fig. 20).

Actin microfilaments are twisted double chains of the protein *actin*. Their diameter is about 7 nm. Microfilaments take part in maintaining and

changing the cell shape, its movements, cell division (in animal cells), and muscle contraction.

Intermediate filaments are so-called because of their diameter (8–12 nm). Their main function is mechanical support. Intermediate filaments are found only in the cells of some animals, including vertebrates. Different types of intermediate filaments are composed of different proteins (e.g., keratin). Intermediate filaments are important for maintaining the shape of the cell and fixing the position of certain organelles. Other intermediate filaments (composed of proteins lamins) form the nuclear lamina, a protein framework that lines the inner part of the nuclear envelope. The outer layer of the epidermis consists of dead cells filled with keratin filaments.

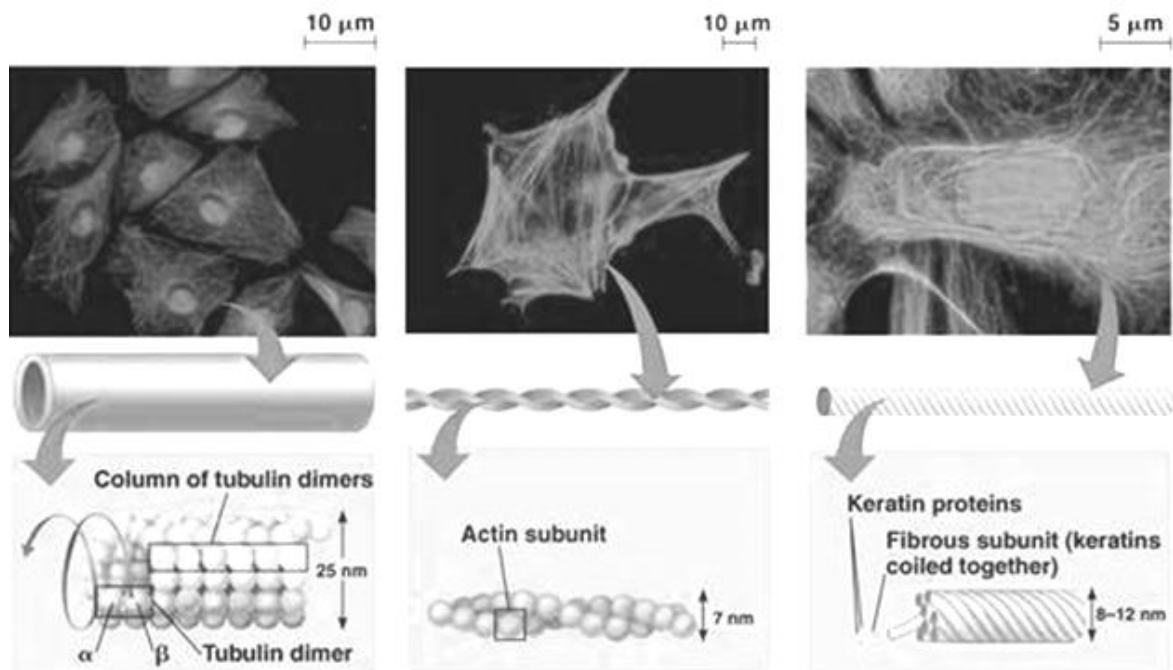


Figure 20. Cytoskeleton elements

Microtubules are hollow rods made of the protein tubulin. Microtubules are constantly disassembled and reassembled in the cell. In animal cells, microtubules “grow” from the centrosome. In addition to their role in maintaining the cell shape, microtubules serve as pathways through which various cell structures are transported. For example, microtubules transport vesicles from the endoplasmic reticulum to the Golgi apparatus. Microtubules are necessary for chromosome separation during mitosis. This transport is performed by motor proteins such as dyneins and kinesins.

Centrioles are barrel-shaped organelles found in animal cells and most protists. Each centriole consists of nine triplets of microtubules (the structure is called 9 + 0). They are arranged in pairs forming 90° angles near the nucleus (Fig. 21).

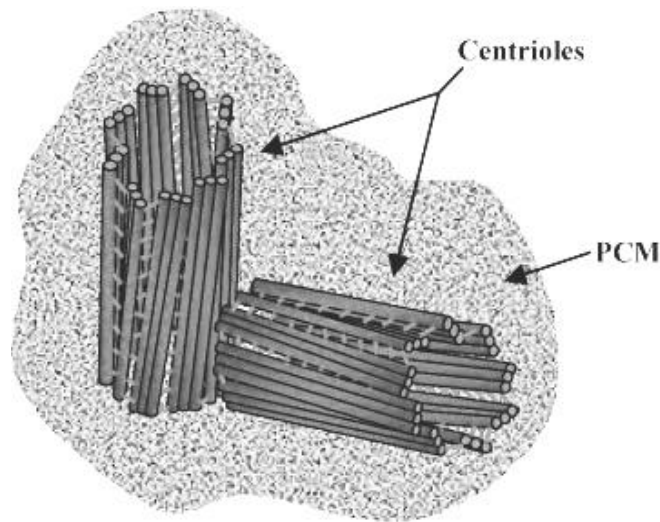


Figure 21. Centrioles and pericentriolar material

The region of the cytoplasm containing a pair of centrioles is called the centrosome.

The pericentriolar material has tubulin and serves as the “start point” of microtubule assembly. Thus, the centrosome is a microtubule organizer. Centrosomes are also responsible for the reorganization of microtubules that occurs during cell division.

Cilia and flagella are thin outer outgrowths of the cytoplasm of some eukaryotic cells. They have a similar structure, but cilia are usually present on the cell surface in large amounts, while flagella are usually fewer in number (one or several). The flagellum or cilia of a eukaryotic cell is covered with the membrane.

Bacterial flagella have a completely different structure. They are not covered by the membrane and are filaments of the flagellin protein. The bacterial flagellum, unlike the eukaryotic one, is not capable of bending. It is spun by a motor in its basement.

Pili (singular: pilus) are protein filaments located on the surface of many bacteria. Their function is to attach prokaryotes to each other and/or to the substrate. Special sex pili are used by bacteria to transfer DNA during bacterial conjugation from the donor cell to the recipient cell.

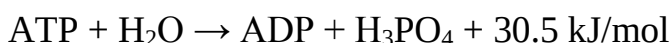
METABOLISM. THE CATABOLIC SYSTEM OF THE CELL

Metabolism is the sum of all chemical reactions in the body. Production of various molecules usually requires not one reaction but a chain of several reactions. Each reaction is catalyzed by a specific enzyme. Such a chain of reactions is called the *metabolic pathway*. Mutation in the genes coding for such enzymes often causes hereditary metabolic disorders.

Metabolism includes the reactions requiring energy, in which complex molecules are synthesized from simpler ones. This part of metabolism is called anabolism (or assimilation). The sum of all reactions in which compounds are broken down into simpler ones and energy is released is called catabolism (or dissimilation). The energy obtained by the cell during catabolic reactions can be used for:

- mechanical work (e.g. movement of cilia, flagella, muscle contraction, movement of chromosomes during cell division);
- transport of substances across the membrane against their concentration gradient;
- construction of chemical bonds (e.g. synthesis of proteins).

The energy source for the proteins performing such work is often hydrolysis of adenosine triphosphate (ATP), and such proteins are called ATP-ases.



ATP is basically a nucleotide consisting of a ribose linked to adenine and a chain of three phosphoric acid residues. When one of the phosphates is cleaved, energy is released and adenosine diphosphate (ADP) is formed. Then ADP can be bound to phosphate again (this requires energy). The energy for such ATP regeneration can be obtained from the chemical bonds of other molecules (e.g., glucose).

ATP production. In a typical eukaryotic cell, most ATP is formed from ADP and phosphate during **oxidative phosphorylation**. This process is performed by the protein complex called **ATP synthase**. It is located on the inner mitochondrial membrane. The energy source for its work is the difference in proton concentrations on either side of the inner membrane: there are more protons in the intermembrane space than in the matrix. Similar to the flow of water providing energy to a water mill, the flow of protons passing through ATP synthase down the electrochemical gradient (i.e. from the intermembrane space to the matrix) provides it with energy. This process is called chemiosmosis.

The high concentration of protons in the intermembrane space is maintained by four types of protein complexes. Together they are called the **electron-transport chain (ETC)**. These complexes are also located in the inner mitochondrial membrane, where they pump protons from the matrix to the intermembrane space. The energy source for ETC is the transfer of electrons from electron-carriers (molecules of NADH, FADH₂) to O₂ (strong attractor of electrons). In ETC, electrons pass through a chain of *redox centers* — special structures capable of capturing an electron to themselves. The redox centers are arranged in such a way that each following one has a higher affinity to the electron than the previous one. As a result, when an

electron transferred from NADH comes to a redox center, it is immediately attracted to the next one, and from it to the next one, and so on. The generated electron current is the source of energy for pumping protons into the intermembrane space. The oxygen molecules which are electron-acceptors are then combined with the protons to form water.

NAD^+ can acquire electrons (i.e., it can be reduced to NADH) through several processes. They are considered in the example of glucose breakdown.

1. *Glycolysis*. Outside mitochondria, in the cytosol, the glucose molecule is broken down into two molecules of pyruvic acid (pyruvate) in a chain of ten reactions. In this process, 2 NAD^+ are reduced and 2 ATP are formed.

2. *Oxidative decarboxylation of pyruvate*. The pyruvate formed during glycolysis is transferred to the mitochondria where the carboxyl group is cleaved from it (forming CO_2). The remaining acetyl group is transferred to coenzyme A, resulting in *acetyl-CoA*. During this process, each pyruvate molecule is used for the reduction of one NAD^+ to NADH.

3. *Citric acid cycle (Krebs cycle)*. Acetyl-CoA enters the cycle of reactions in which it is completely broken down, reducing 3 NAD^+ , 1 FAD, and forming 1 molecule of GTP.

All the NADH and FADH_2 are enough to produce 28 ATP. With the 2 ATP and 2 GTP produced during glycolysis and the Krebs cycle, the final number of **ATP produced from one glucose molecule is up to 32**.

Diseases associated with dysfunctions of lysosomes and peroxisomes. Lysosomal storage diseases are a group of hereditary metabolic disorders in which various substances that are normally broken down in the lysosomes accumulate because of deficiency or lack of specific enzymes. Lysosomal storage diseases have a progressive course and, depending on the specific disease and subtype, can be fatal in early childhood. Examples: Hurler syndrome, Tay–Sachs disease, Gaucher disease, Niemann–Pick disease, Cystinosis, Pompe disease. Examples of peroxisomal diseases: X-linked adrenoleukodystrophy, Zellweger syndrome.

CHAPTER 3

STRUCTURE OF DNA

The structure of DNA was discovered by J. Watson and F. Crick in 1953, based on data obtained by E. Chargaff, R. Franklin, and M. Wilkins. DNA (deoxyribonucleic acid) consists of a large number of nucleotides. Each DNA nucleotide consists of three components (Fig. 22):

- five-carbon sugar (i.e., pentose) called deoxyribose;
- phosphoric acid residue (i.e., phosphate);
- nitrogenous base attached to the sugar.

There are four types of DNA nitrogenous bases (Fig. 23): adenine (A), guanine (G), thymine (T), and cytosine (C). The first two (A and G) are purine bases, i.e., they are modified versions of the purine molecule. Thymine and cytosine are pyrimidine bases, i.e., their basis is a molecule called pyrimidine. The nitrogenous bases are attached to the first carbon (1') of the deoxyribose molecule, and the phosphate is attached to the fifth one (5').

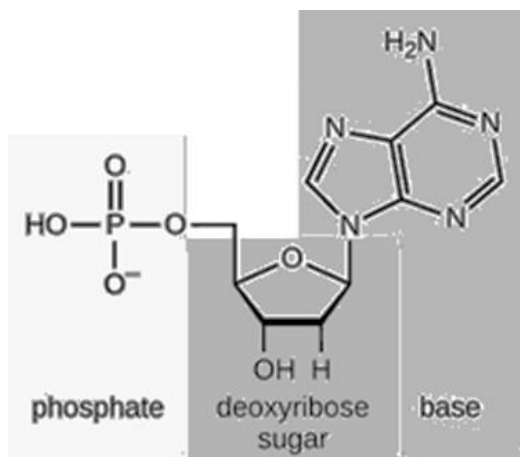


Figure 22. Nitrogenous bases

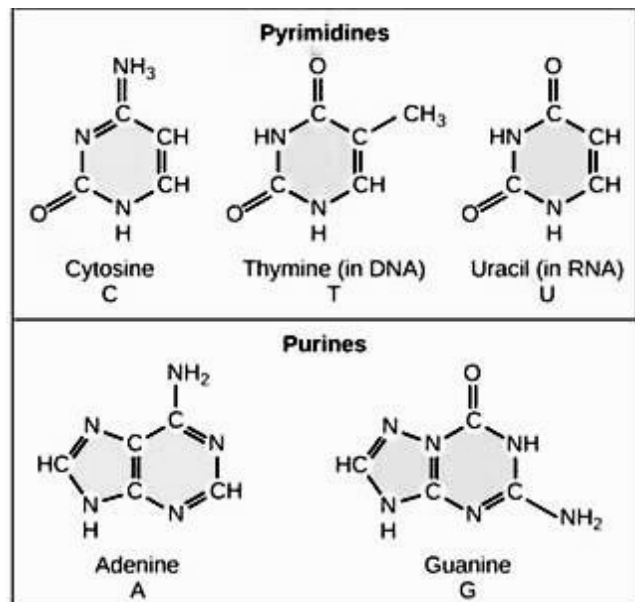


Figure 23. The structure of a nucleotide

Nucleotides can be assembled into a chain. DNA consists of two such chains (strands). The two strands form a double helix (Fig. 24). One turn of the DNA double helix contains about 10 pairs of nucleotides and is about 3.4 nm long. Each nucleotide of a strand is connected with adjacent nucleotides by its fifth (5') and third (3') carbon atoms of deoxyribose. There is always a phosphoric acid residue between the 5' carbon atom of one nucleotide and the 3' of the adjacent one. This bond between the two nucleotides is called a phosphodiester bond (conventionally "C-O-P-O-C").

The basis of a DNA strand is a sugar-phosphate backbone. It consists of multiple repeats of phosphate and deoxyribose. One end of the backbone has the nucleotide that is not bound to the other nucleotides by its 5' carbon atom. The other end of the strand has the nucleotide that is not bound to the others by its 3' carbon. Thus, two ends of a nucleotide chain are called the 5' and the 3' end. The two strands of DNA are always antiparallel — the 5' end of one strand is situated near the 3' end of the other one.

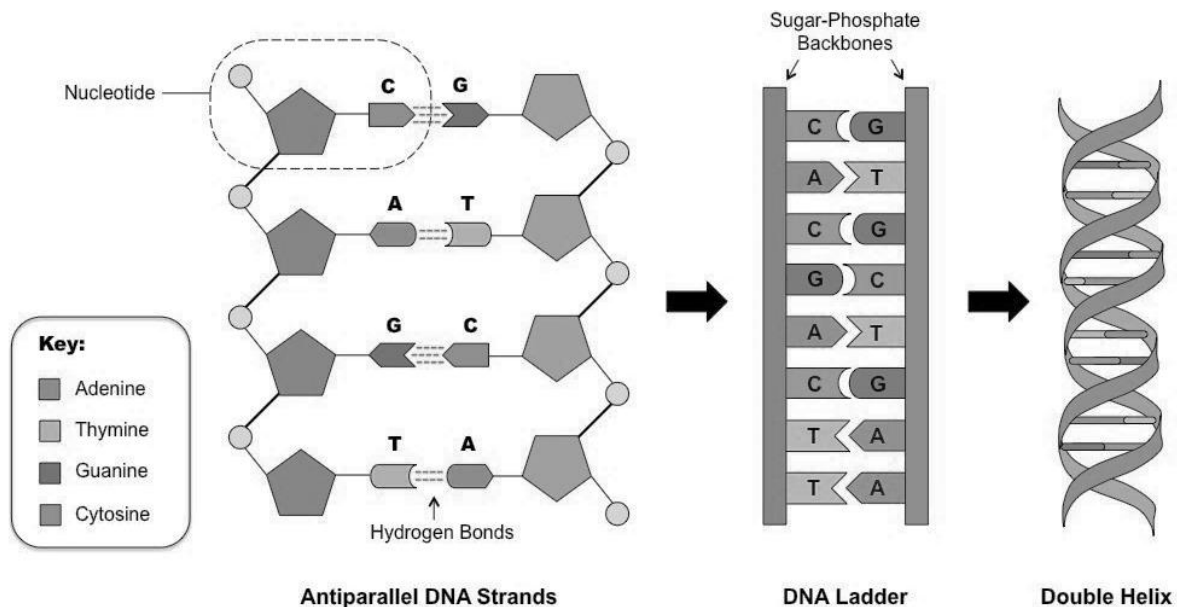


Figure 24. DNA structure

Each deoxyribose of the sugar-phosphate backbone is connected to a nitrogenous base (1 of 4 types), so the nucleotide chain can be considered as a “text” written in a “language” of four letters, which allows a DNA molecule to be a carrier of information.

The connections between the bases of the two nucleotide chains are not random: the purine base is always bound to the pyrimidine base, namely, A is bound to T, and G is bound to C. As a consequence:

1. The amount of adenine in DNA is equal to the amount of thymine (A=T).
2. The amount of guanine in DNA is equal to the amount of cytosine (G=C).
3. The number of purine bases (A and G) in DNA is equal to the number of pyrimidine bases (C and T).

These patterns are known as Chargaff rules, named after the biochemist who discovered them.

EVIDENCE FOR THE ROLE OF DNA IN THE TRANSMISSION OF HEREDITARY INFORMATION

In 1928, the British physician Frederick Griffith studied the bacterium that causes pneumonia in mammals (*Streptococcus pneumoniae*). He worked with two strains (types) of the bacterium: one pathogenic (causing the disease) and one nonpathogenic (harmless). Griffith discovered an unusual fact: if dead pathogenic bacteria (killed by heat) mixed with alive nonpathogenic bacteria, they cause disease in mice. Griffith concluded that some chemical component of dead pathogenic cells was transferred to living cells and changed their properties. But the nature of this substance was not known. Griffith called this phenomenon **transformation**, which is now defined as a change in a bacterium's genotype and phenotype after taking up foreign DNA by the cell.

In 1944, Oswald Avery, MacLeod McCarty and Colin MacLeod identified the transforming substance as DNA: the addition of enzymes that destroy proteins and RNA (proteases and ribonucleases) to preparations derived from bacteria did not disturb the transformation, while processing with deoxyribonuclease (a DNA-digesting enzyme) did.

Other evidence for the role of DNA in the transmission of hereditary information came from the study of **transduction**. It was described by N. D. Zinder and J. Lederberg in 1952. Transduction is the transfer of genes from one bacterial cell to another with the help of specific bacteriophages. Bacteriophages (phages) are viruses that infect bacteria. They inject their DNA into the cell, and the infected cell begins to produce components of new phages. The new phages contain a copy of the original phage DNA. But sometimes a new phage may contain a fragment of bacterial DNA. A cell infected with such phage can acquire new properties.

In 1952, Alfred Hershey and Martha Chase conducted experiments showing that DNA is the genetic material of a phage known as T2. This phage infects the bacterium *Escherichia coli* (*E. coli*). It was known that T2 consists of DNA and protein. In the experiment, Hershey and Chase used radioactive sulfur to label viral proteins and radioactive phosphorus to label its DNA. Some *E. coli* cells were infected with T2 having radioactively labeled proteins and others — with T2 having labeled DNA. The researchers found that the phage DNA had entered the infected cells, but the phage protein had not. The mechanism of genetic information storage in DNA became clear after the structure of DNA was decoded.

Definition of a gene:

1860–1900: “A gene is a discrete unit of heredity.”

The 1910s: “A gene is a separate locus of the chromosome.”

The 1940s: “A gene is a protein circuit.”

The 1950s: “A gene is a physical molecule.”

The 1960s: “A gene is a transcribed code.”

The 1970s–1980s: “A gene is an open reading frame for a part of a DNA sequence.”

1990–2000: “A gene is a structural unit of the genome registered in a database.”

2007: “A gene is a subprogram in the operating system of the genome.”

2009: “A gene is an association of genomic sequences encoding a coherent set of potentially overlapping functional products as part of genetic regulatory networks.”

Viral DNA and RNA. The genomes of various viruses contain RNA or DNA, which can be single-stranded or double-stranded. The DNA or RNA can be **linear** or **circular** (except for double-stranded RNA which is always linear). The genome can be segmented into several molecules. The RNA of some can be translated immediately (**plus-strand**). In others, a complementary RNA strand must be created for translation (the RNA of such viruses is called **minus-strand**). The number of genes in a genome is usually several to several thousand (in eukaryotes, thousands).

DNA in bacteria. In bacteria, genetic material is the circular DNA molecule of a bacterial **chromosome** (nucleoid) and multiple small circular DNA called **plasmids**. The **nucleoid** is a region in the cytoplasm of a bacterial cell where the bacterial chromosome is located. It is a densely packed circular DNA bound to histone-like proteins. The size of bacterial DNA is usually several million pairs of nitrogenous bases. For example, the *E. coli* genome is 4.6 million bp. In addition, bacteria also commonly have several small circular DNAs that replicate independently. They can range in size from a few thousand to hundreds of thousands of base pairs, and the number of copies per cell can range from one to several hundred. Plasmids usually contain only a few genes, which can be useful under certain conditions, for example, to give resistance to particular antibiotics.

Levels of hereditary material organization:

1. The structural unit at the **gene level** of genetic material is a gene. As genes are separate units, they can be inherited independently of one another (3rd Mendel’s law) and independent changes (mutations which lead to changes in the characters) in the genes are possible.

2. Genes of eukaryotes are in chromosomes — **chromosome level** of genetic material. The genes of each chromosome (pair of homologous chromosomes) comprise linkage groups and are inherited together. At this level, linkage and recombination (combining the parental genes during sexual reproduction due to random segregation of chromatids and crossing over) of genes occur.

3. All the genes of the organism are functionally integrated and comprise a system — genotype (or genome). The same gene can have a different effect on different genotypes. The **genome level** explains intra- and inter-allelic interactions of genes of a chromosome or even several chromosomes.

DNA ORGANIZATION IN EUKARYOTES

The genome of eukaryotes is contained in several (or many) chromosomes located in the nucleus.

Nuclear DNA is bound with proteins (histones and non-histone proteins). The complex of DNA and proteins is called **chromatin**.

Histones are proteins that provide DNA condensation (i.e., its tight packing) and participate in the regulation of gene expression (i.e., in “turning on” and “turning off” the transcription of genes). These proteins easily interact with negatively charged DNA, since they have a large number of positively charged amino acids.

Four types of histones (two of each type, **H2A**, **H2B**, **H3**, and **H4**) form *octamers*, i.e., structures consisting of eight polypeptides.

DNA makes approximately 1.65 turns around each octamer (146 base pairs), and in combination with another histone, **H1**, almost two turns. The formed complexes of histone octamers and DNA are called **nucleosomes** (Fig. 25).

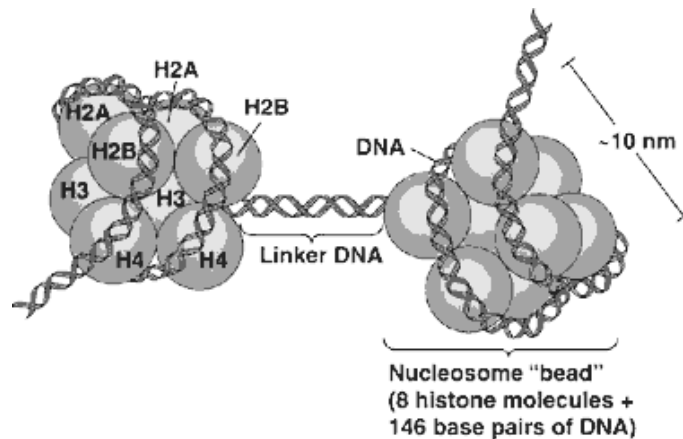


Figure 25. Diagram of nucleosomes and DNA wrapped around them

Histone H1 is not involved in the formation of the histone octamer. It fixes DNA and participates in the denser packing of chromatin. The described mechanism makes DNA 7-fold shorter.

Histone-bound DNA is often compared to “**beads on a string**”, where the “beads” are nucleosomes and the “string” DNA between them (it is called **linker DNA**). This structure is called a **10-nm fiber** because of its diameter (Fig. 26). In the laboratory, it has been shown that, under certain conditions, it forms a more densely packed **30-nm fiber**, also called a solenoid. Recent studies revealed that 10-nm fiber is the main component of chromatin. Though 30-nm fiber may exist in some cells in some conditions. This is the area of active research.

Sections of DNA in the nucleus are coiled into small globules, which, in turn, are coiled into bigger globules. The globules containing between 0.1 and 1.5 million base pairs are called **topologically associated domains (TADs)**. In addition, many DNA sites form **loops**.

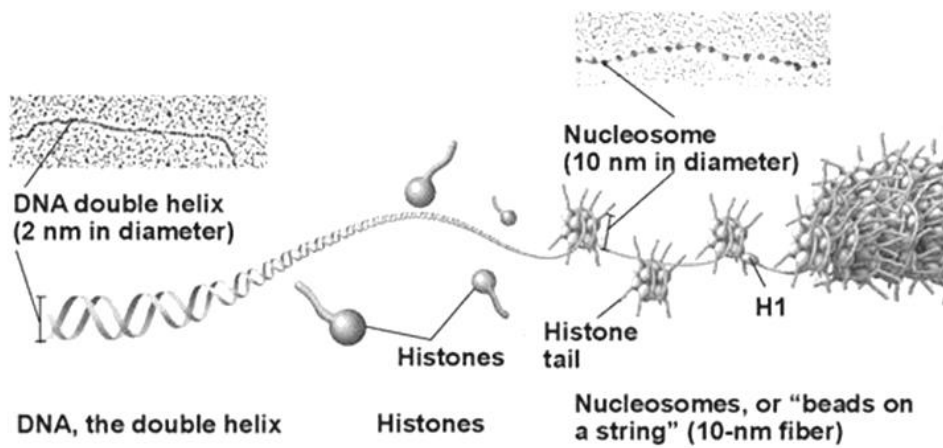


Figure 26. Chromatin organization

During cell division, chromatin is organized into chromosomes:

- First, the protein **condensin II** binds to the 10-nm fiber and forms loops. It also forms a *central scaffold*, from which the loops continue to extrude over time. As the loops grow, the chromosome becomes wider and shorter.
- **Condensin I** then binds to DNA outside the central scaffold, forming smaller loops from larger ones (Fig. 27). The process continues, with more and more loops coming out and the chromosome becomes denser, shorter, and wider.
- In addition, the scaffold itself also begins to twist into a helix. In metaphase, the chromosome is in its densest state.

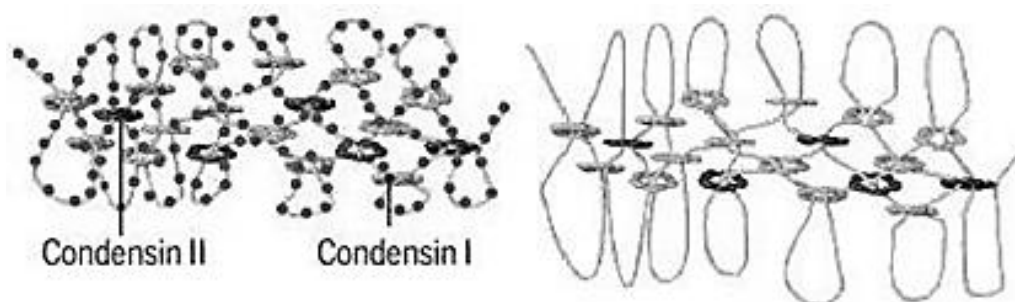


Figure 27. Formation of the chromosome from 10-nm fiber

THE STRUCTURE OF THE NUCLEUS

The nucleus is the compartment of a eukaryotic cell that contains its genetic information (Fig. 28).

Its average size is usually about 5 μm .

The **nuclear envelope** consists of **inner and outer membranes**. Each of them is a lipid bilayer. The space between the membranes is called **perinuclear** (or **intermembrane**) **space**. It is a single compartment with granular ER surrounding the nucleus.

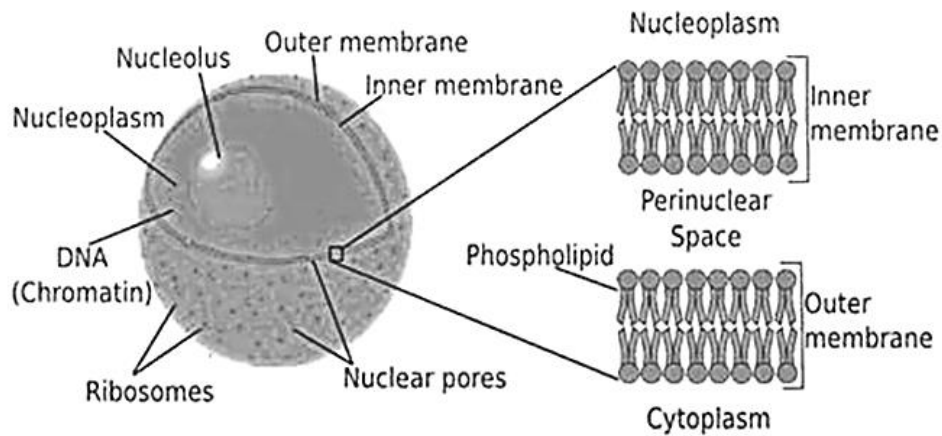


Figure 28. The structure of the nucleus

There are multiple pores in the nuclear envelope. In each pore, the two membranes of the nucleus merge to form a hole about 80 nm in diameter. It contains a complex structure of nucleoporin proteins controlling transport through the pore — the **nuclear pore complex** (Fig. 29). Its structure includes nuclear and cytoplasmic rings (facing the nucleus and cytoplasm, respectively). Cytoplasmic filaments extend from the cytoplasmic ring into the cytosol. The nuclear ring has nuclear filaments, the ends of which are connected to the terminal ring. As a result, a nuclear basket is formed.

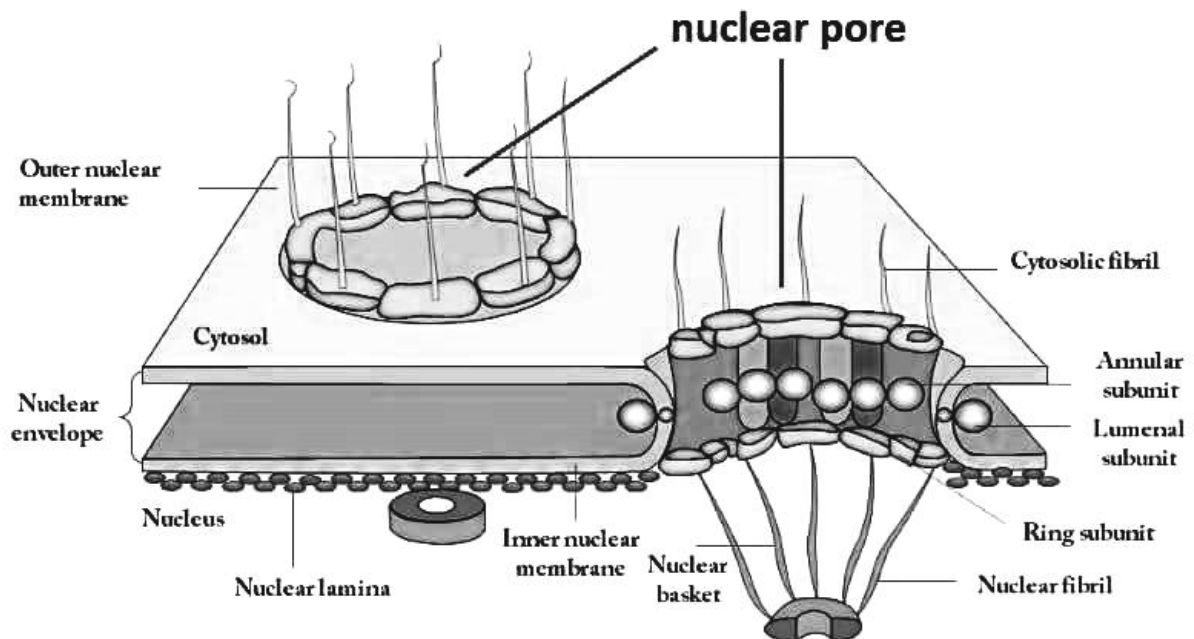


Figure 29. Nuclear pore complex

There is a channel with a diameter of about 40 nm in the pore complex. Relatively small molecules can pass through the pores freely.

Large macromolecules are transported with the participation of special proteins. For example, proteins designated for import into the nucleus must

have a certain short amino acid sequence, which is called a **nuclear localization signal** (NLS), and for export from the nucleus a nuclear export signal (NES).

The inner surface of the nucleus is lined with a network called the **nuclear lamina** consisting of a type of intermediate filaments. In animals, they consist of **lamin** proteins.

There is evidence that the nucleus has a network of proteins called the **nuclear matrix**.

The lamina and matrix interact with chromatin and take part in its spatial arrangement in the nucleus and also relate to the regulation of gene function.

The nucleus contains one or more **nucleoli** (singular is nucleolus).

A nucleolus is a dense mass of granules and fibrils adjacent to chromatin. The nucleoli are the place where ribosomal RNA is produced.

Ribosomal proteins arrive here from the cytoplasm. They create a complex with rRNA and form the precursors of the large and small ribosome subunits.

They are further exported from the nucleus through the nuclear pores into the cytoplasm.

Nucleoli form in specific sites of some chromosomes. These sites are called the nucleolar **organizer regions** (NORs). They contain multiple copies of genes coding for ribosomal RNAs (3 out of 4 rRNAs).

Nucleolus consists of the fibrillar center (containing DNA of NORs), the dense fibrillar component, which contains synthesized rRNAs, and the granular component, where the ribosome subunit precursors are assembled.

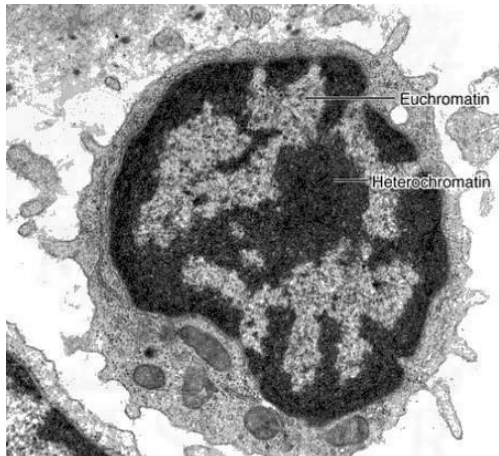


Figure 30. TEM of the nucleus with chromatin

The bulk of the nucleus is filled with chromatin. The electron microscope shows that chromatin is heterogeneous: among the lighter mass of less densely packed chromatin (called **euchromatin**), there are darker regions of more densely packed chromatin (i.e., **heterochromatin**, Fig. 30). As previously described, the structural unit of both types of chromatin is a 10-nm fiber. In heterochromatin, it is organized more tightly and the enzymes have no access to genes — the genes of heterochromatin are silenced. Modification of the chromatin structure that changes the access of

regulatory proteins to genes is called **chromatin remodeling**. It is performed either by covalent modification of histones with specific enzymes or by displacement or removal of nucleosomes.

In the past, it was thought that the chromosomes in the nucleus of a dividing cell were evenly mixed. But in fact, each chromosome fills a particular area in the nucleus.

THE STRUCTURE OF CHROMOSOMES

In metaphase, each chromosome consists of two “halves” called **chromatids** (Fig. 31). Each of the chromatids contains one of two copies of the same DNA made during DNA replication. This is necessary for the precise transfer of hereditary information to the next generation: the chromatids of each chromosome of the mother cell will separate and become the chromosomes of the daughter cells.

Connections between two **sister chromatids** (i.e., chromatids of the same chromosome) are provided by proteins **cohesins**.

Each sister chromatid has a **centromere**, the region in which chromatids are most closely linked to each other. This close connection is provided by proteins that recognize the centromeric DNA (it contains distinctive repetitive sequences) and bind to it. Other proteins provide a tight packing of DNA in this region, forming a constriction typical of the centromere. Due to this constriction, the centromere is also referred to as a **primary constriction**. Metaphase chromosome centromeres also bind to **kinetochores** — protein complexes that connect chromatids to microtubules during cell division.

The parts of the chromatid on either side of the centromere are called **arms**. Since the centromere is often not in the center of the chromatid, one arm is **short (p-arm)** and the other is **long (q-arm)**.

Therefore, all human chromosomes can be divided into **metacentric** (the arms are almost equal), **submetacentric** (the arms are markedly different), and **acrocentric** (one arm is much shorter than the other).

In human chromosomes, the short arms of chromosomes 13, 14, 15, 21, and 22 have **secondary constriction**. The DNA sites in these regions contain ribosomal RNA genes, i.e., they are **nucleolar organizer regions**.

The part of the chromosome arm separated from the rest of the chromatid by the secondary constriction is called a **satellite**, and chromosomes containing them are called **satellite chromosomes**.

The terminal regions of each chromatid are called **telomeres**. They contain the terminal sites of the DNA encapsulated in the chromatid. In vertebrates, telomeres contain a repetitive TTAGGG sequence. The telomere length in human chromosomes is between 5,000 and 15,000 bp. The telomere DNA forms loops that are stabilized by special proteins. This helps to hide the terminal DNA sites from the cell systems responsible for repairing breaks in the DNA.

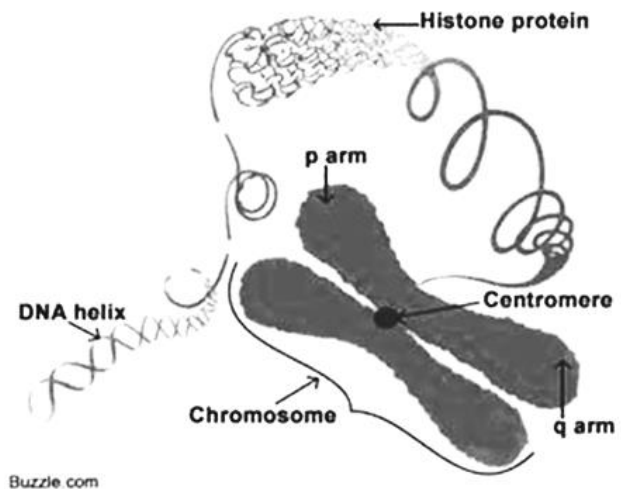


Figure 31. Chromosome structure

Some regularities common to the chromosomes of different biological species are called chromosome rules:

The rule of constant number. Somatic cells of each species normally have a strictly defined number of chromosomes; for example, in humans, it is 46, and in *Drosophila* — 8 (but there are some exceptions).

The rule of pairing. Each chromosome in a diploid set has a homologous chromosome — similar in size and centromere location. One of these chromosomes comes from the mother and the other from the father. The same loci (i.e., sites) of homologous chromosomes contain the same genes (however, it should be noted that the same gene can have several variants — alleles — and they can differ).

The rule of individuality. Each pair of chromosomes differs from the other pair in size, centromere location, and the set of genes it contains.

The rule of chromosome continuity. In the process of genetic material doubling, a new DNA molecule is synthesized based on the information of the old DNA molecule (matrix synthesis reaction), and, therefore, each daughter chromosome is derived from the mother chromosome.

It should be noted, these are not strict rules, but rather general observable patterns. If the number of chromosomes could not change, then it would have remained constant during evolution. Traces of such changes are also known for humans.

KARYOTYPE, IDIOGRAM. CLASSIFICATION OF CHROMOSOMES

The set of features of all chromosomes that are specific to a cell is called a **karyotype** (Fig. 32). The karyotype reflects the number of chromosomes, their shape, size, and characterizes an individual or biological species. Chromosomes can be studied by microscopy of specially prepared cell samples.

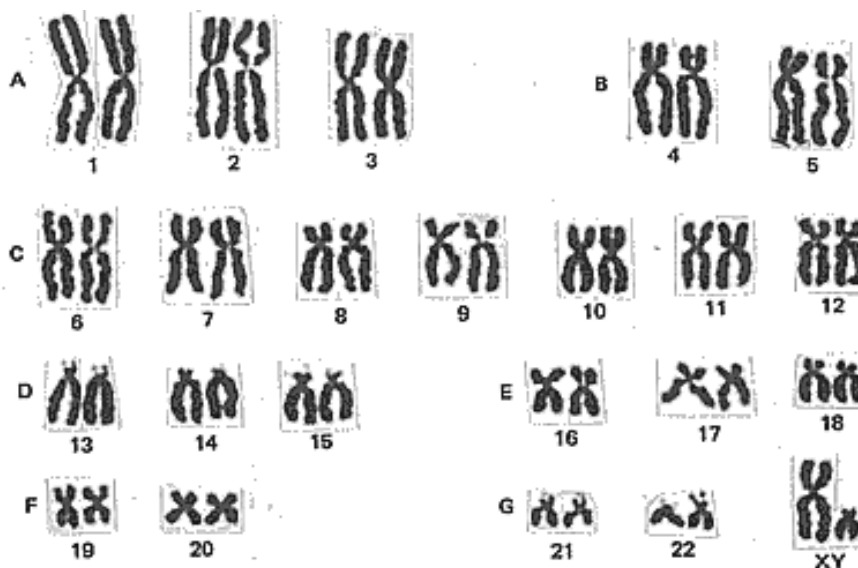


Figure 32. Karyotype of human

An image in which chromosomes photographed under a microscope are divided into groups and sorted by decreasing size and centromere position is called a **karyogram**. A diagram in which chromosomes are also sorted by size and centromere position and the differential staining is shown is called an **idiogram**. It often shows one chromosome from each pair.

A diagram of a single differentially stained chromosome is also called an **ideogram** (Fig. 33).

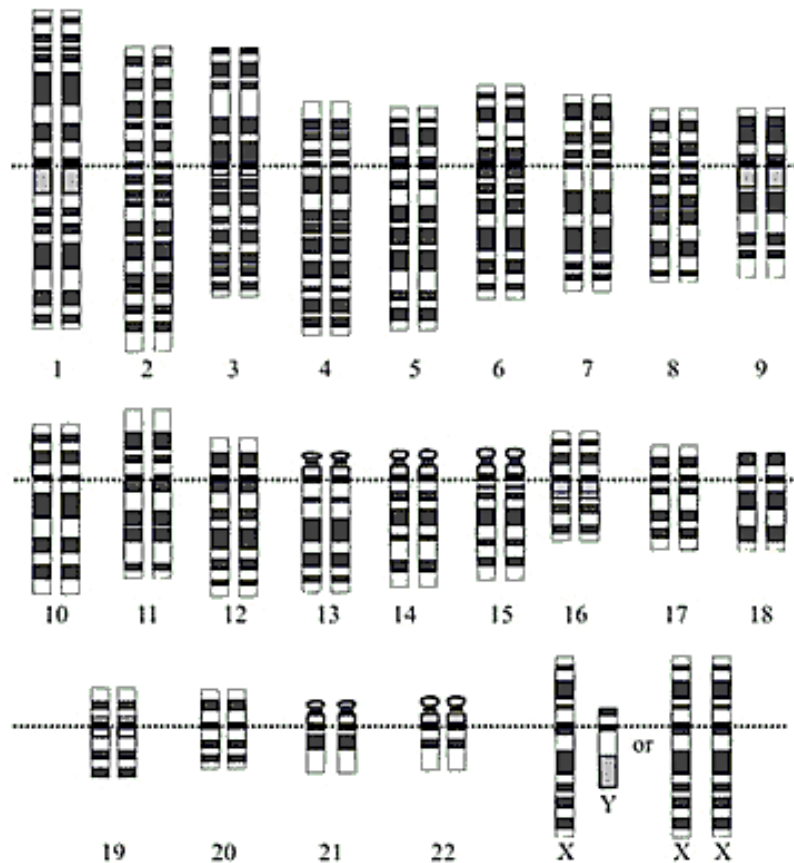


Figure 33. Human ideogram

The karyotype of a human somatic cell contains 46 chromosomes, of which 23 are derived from the mother and another 23 from the father. Therefore, humans have 23 pairs of chromosomes. It should be noted that we are talking about somatic cells, i.e. the cells that form the human body, and not the sex cells, which are used for sexual reproduction. Somatic cells have a double set of chromosomes and are called diploid ($2n$). Sex cells have a single set of chromosomes and are called haploid ($1n$).

One pair is called the **sex chromosomes**, or heterochromosomes (X and Y in men, X and X in women).

The chromosomes of the other 22 pairs are the same in both males and females and are called **autosomes**.

The male karyotype is written as “46, XY” or “44+XY”, and the female karyotype is “46, XX” or “44+XX.”

The first common classification of human chromosomes was the **Denver classification**, 1960 (Table 2). According to this classification, all chromosomes were divided into 7 groups (from **A** to **G**) according to their size and the centromeric index. The **centromeric index** (CI) is the percentage ratio of the short arm length to the length of the entire chromosome. Chromosomes were arranged in descending order of length and then in descending order of CI. However, it was not possible to distinguish chromosomes of the same group from each other.

Table 2

Denver classification

Chromosome	Group	Morphological features
№ 1–3	Group A	These are large, metacentric and submetacentric chromosomes, with CI 38–49. The first pair of chromosomes is the largest metacentric (CI 48–49), and there may be a secondary constriction in the proximal part of the long arm near the centromere. The second pair of chromosomes is the largest submetacentric pair (CI 38–40). The third pair of chromosomes is 20 % shorter than the first, submetacentric (CI 45–46), easily identified
№ 4–5	Group B	These are large submetacentric chromosomes; CI 24–30. They are not differentiated from each other by normal staining. The composition of R- and G-segments (see below) is different
№ 6–12, X	Group C	The chromosomes are medium-sized, submetacentric, CI 27–35. The 9 th chromosome often has a secondary constriction. The X chromosome also belongs to this group. All chromosomes in this group can be identified by Q- and G-staining
№ 13–15	Group D	The chromosomes are acrocentric, very different from all other human chromosomes; CI is about 15. All three pairs have satellites. The long arms of these chromosomes differ in Q- and G-segments
№ 16–18	Group E	The chromosomes are relatively short, metacentric or submetacentric, CI 26–40 (chromosome 16 has a CI of about 40, chromosome 17 has 34, chromosome 18 has 26). In the long arm of chromosome 16, a secondary constriction is detected in 10 % of cases
№19-№20	Group F	Chromosomes are short, submetacentric, CI 36–46. In normal staining, they appear identical, but in differential staining, they are well distinguishable
№ 21–22, Y	Group G	The chromosomes are small, acrocentric, CI 13–33. The Y chromosome also belongs to this group. The chromosomes are easily distinguishable by differential staining

In 1971, the **Paris nomenclature** of human chromosomes was proposed. It was based on the differential staining of chromosomes, resulting in the formation of light and dark bands on them (see the chapter “Human genetics”). The location and size of the resulting bands are not random but unique to each of the human chromosomes. This makes it possible to recognize each chromosome and not only determine its group according to the Denver classification. In addition, differential staining opens up the possibility of indicating the exact location of a particular gene. For example, the *HBB* gene encoding the hemoglobin beta chain is located in the short arm of the eleventh chromosome (11p). In this arm, we distinguish regions, in regions are bands, and further on are sub-bands. The *HBB* gene is located in the fourth sub-band (.4) of the fifth band (5) of the first region (1). Thus, the location of this gene can be designated as: 11p15.4.

CYTOPLASMIC INHERITANCE

In addition to nuclear DNA, there is some DNA in the cytoplasm. Genes of such DNA are called **plasmagenes**. Examples of such DNA are mitochondrial (**mtDNA**) and plastid DNA. The presence of DNA in mitochondria and plastids is associated with their descent from bacteria. There may also be foreign elements in the cytoplasm, such as viruses and bacteria. Inheritance of traits controlled by genes located in the cytoplasm of the cell rather than in the chromosomes in the cell nucleus is called extranuclear inheritance (cytoplasmic inheritance).

The criteria for cytoplasmic inheritance are:

- Maternal inheritance — from the mother, but not the father, to all offspring.
- Absence of quantitative Mendelian segregation in offspring.
- Absence of genetic linkage.
- Different results of reciprocal crosses.

PLASTID INHERITANCE

In 1908, Carl Correns discovered a species of the marvel of Peru (four o'clock flower, *Mirabilis jalapa*) in which some branches had white leaves, some had green leaves, and some had variegated leaves. The white leaves and white patches of variegated leaves lacked chlorophyll (because of a mutation in chloroplast DNA). In his experiments, Correns found that if seeds were obtained from buds on branches with green leaves, all progeny plants carried only green leaves, regardless of the pollen source phenotype. Correns concluded that the trait is transmitted through the cytoplasm of the maternal germ cell.

MITOCHONDRIAL INHERITANCE

In 1952, Mary B. Mitchell and Herschel K. Mitchell studied the bread mold *Neurospora crassa*. They discovered a slow-growing mutant strain. The slow growth was associated with impaired mitochondrial function caused by the absence of several cytochrome proteins necessary for electron transfer in the electron-transport chain. Another example was described by Boris Efrussi and his colleagues in 1956 in baker's yeast (*Saccharomyces cerevisiae*). He found that yeast had petite (small) colonies that grew slowly because of mutations in mtDNA.

Maternal inheritance of mtDNA is also characteristic of some human diseases. Typically, mtDNA in animals has 16–18,000 base pairs; in vertebrates, there are five to ten such molecules per organelle. Human mtDNA contains 16,569 base pairs and contains 37 genes encoding 13 proteins, 22 tRNAs, and 2 rRNAs. However, most of the proteins that function in mitochondria are encoded by nuclear genes.

Mitochondrial DNA is especially vulnerable to mutations. One of the reasons is that mitochondria produce large amounts of reactive oxygen species during cellular respiration. They damage various molecules. As a result, mutations in mtDNA occur about ten times more frequently than in nuclear DNA.

Individual cells can have several thousand mitochondria. A zygote receives a large number of organelles through the oocyte, so if the mutation is only in one or a few organelles, it may not affect the cell. During embryonic development, cell division disperses the initial mitochondrial population of the zygote, and these organelles multiply autonomously in the forming cells. Therefore, an embryo and then an adult organism will have cells with normal and mutant organelles. The presence of more than one type of mtDNA in an individual is called heteroplasmy. Mitochondrial diseases are characterized by a correlation between the severity of the course and the number of mutant forms of mtDNA. More than 150 human syndromes are currently known to be inherited in this way. They mostly affect tissues that need a large amount of energy (skeletal muscles, brain).

Examples of *mitochondrial diseases*:

- MERRF syndrome (myoclonus epilepsy with ragged red fibers). It is a maternally inherited disease. MERRF syndrome is manifested by muscle coordination disorders, deafness, dementia, and epileptic seizures. Skeletal muscle structure abnormalities are observed, and brain function is impaired. The cells of sick people have a mixture of normal and abnormal mitochondria in different proportions in different patients. Apparently, the presence of only mutant mitochondria is fatal.

- Hereditary Leber Optic Neuropathy (LHON) is maternally inherited and is characterized by sudden bilateral blindness. The average age of vision loss is 27 years, but the onset is highly variable. In more than 50 percent of cases, the disease is caused by a mutation in the mtDNA gene encoding the NADH dehydrogenase subunit.

- Kearns-Sayre syndrome (KSS) is characterized by loss of vision, hearing, and heart abnormalities. The genetic basis of the syndrome includes deletions at different mtDNA sites. Many patients develop symptoms in adulthood. The proportion of mutant mtDNA detected is related to the severity of the symptoms.

In addition to mitochondrial diseases, mitochondrial dysfunction is associated with a large number of diseases, including atherosclerosis, Alzheimer's disease, schizophrenia, various types of cancer, and many others. A link between mitochondrial gene damage and aging has been established.

It is estimated that 1 in 5,000 people has mitochondrial DNA abnormalities or is at risk of developing them. Currently, there is no effective treatment for mitochondrial diseases, but an approach to prevent them is being developed: a cytoplasmic transplant from the female donor's cells into the mother's ovum during in vitro fertilization.

INHERITANCE OF PARASITES AND ENDOSYMBIONTS

Foreign genetic elements can also be inherited with the cytoplasm. For example, the infusoria *Paramecium aurelia* has cell lines that secrete a toxin that is harmless to the cells that produce it but kills other cells of the same species. This is because their cytoplasm contains so-called kappa particles, which represent endosymbiotic bacteria whose preservation in the cytoplasm also depends on nuclear genes. Another example is *Wolbachia* bacteria which live in insect cells and are inherited through the egg cytoplasm. In pictures of such eggs, one can see how the bacteria are concentrated in the area of the egg from which the genital cells will develop, which increases their chance of being passed on to the next generation.

CHAPTER 4 THE CELL CYCLE

THE CELL CYCLE

According to cell theory, new cells emerge only from already existing ones by means of division. **Mitosis** is the most common type of cell division in eukaryotes. The period between divisions is called the interphase. The sequence of all the processes carried out by the cell, leading to its reproduction, is called the **cell cycle** (Fig. 34). It includes the entire life cycle of a cell from its emergence to its division into daughter cells — i.e., **interphase** and division itself. **Interphase** is also divided into three more phases: **G₁** (gap 1 phase, or growth 1 phase), **S** (synthesis), and **G₂** (gap 2 phase, or growth 2 phase). The G₁ phase is also called postmitotic or presynthetic, and G₂ is called premitotic or postsynthetic. So, the cell cycle is an alternation of four phases: **G₁ → S → G₂ → M**.

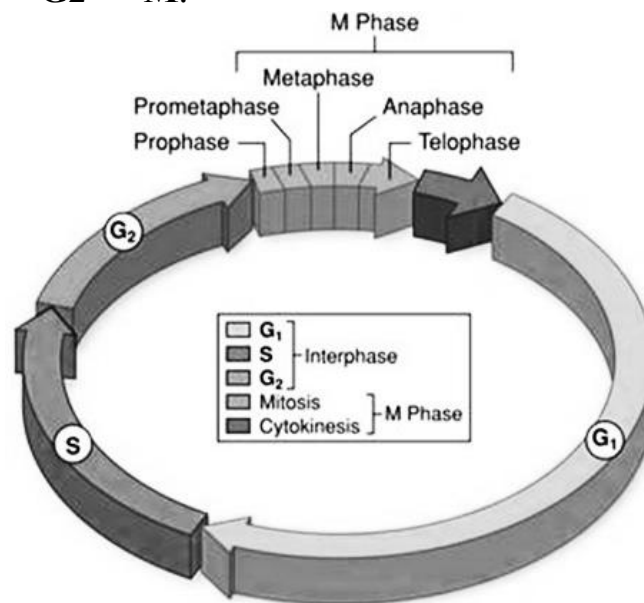


Figure 34. Cell cycle phases

Since both daughter cells must receive complete genetic material during division, all nuclear DNA is first doubled in S-phase and then divided between the two daughter cells in the anaphase of mitosis. Therefore, the content of genetic material from anaphase to S-phase is $2n1chr2s$ (the cell is diploid, and each chromosome contains one chromatid, the cell has two sets of DNA molecules), and after S-phase and before the next anaphase is $2n2chr4c$ (the cell is diploid, each chromosome contains two chromatids and the cell has four sets of DNA).

The duration of the cell cycle and the phases can vary significantly in different cell types. A typical human cell capable of mitosis can divide approximately every 24 hours. The M-phase takes less than 1 hour, the S-phase

takes 10–12 hours, the G2 phase usually lasts 4–6 hours, and the rest of the time is G1. The duration of the G1 phase varies the most in different cell types.

Interphase:

1. The **G1 phase** begins immediately after the completion of mitosis and ends with the beginning of the replication of nuclear DNA. The content of genetic material is $2n1chr2c$. At this time the cell starts to grow, and intensively produces mRNA and proteins. The cell's passage through this phase is influenced by the availability of nutrients and signals from the external environment, such as **growth factors** (described later). When the cell reaches the required size and synthesizes the necessary proteins, S-phase can begin. Not all cells undergo divisions. Most cells enter a dormant stage called **G0** and can stay there for a long time. Others, such as liver cells, may return to the G1 phase and divide in response to different signals (e.g. related to tissue damage).

2. During the **S phase**, DNA replication (i.e., its doubling) occurs. As a result, the content of genetic material in the cell becomes $2n2chr4c$ and remains until the anaphase of mitosis.

3. The **G2 phase** takes place after replication. During this phase, centriole doubling occurs, as well as active tubulin synthesis (microtubules take part in the division of sister chromatids between daughter cells). By the end of the G2 phase, the cell volume increases approximately twice as much as at the beginning of the interphase. Chromatin condensation begins, but the outlines of the chromosomes are not yet visible. As soon as chromosome condensation becomes sufficient to distinguish individual chromosomes with a light microscope, we can conditionally state that the first stage of mitosis — prophase — begins.

DNA REPLICATION

All cells multiply by division. Since both daughter cells must contain equal copies of the mother cell's genome, the cell's DNA is doubled before dividing. This process is called replication.

During replication, DNA strands are separated, and new strands are added to the nucleotides of each strand according to the principle of complementarity. This creates two identical DNA molecules, each consisting of two strands — one “old” and one “new.” This method of DNA doubling is called semi-conservative synthesis (Fig. 35). The fact that replication occurs in this way was confirmed by M. Meselson and F. Stahl in 1958.

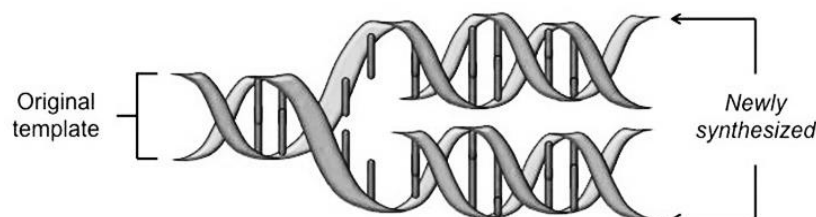


Figure 35. Semi-conservative principle of DNA replication

The DNA replication process is similar in all living organisms. The main role is played by DNA polymerase enzymes, which can add new nucleotides to the existing nucleotide chain according to the principle of complementarity. Some difficulties in presenting this process are related to the fact that:

- First, DNA polymerases cannot initiate replication but can only add nucleotides to a pre-existing fragment of a new strand. Therefore, such fragments — primers — are specially synthesized by the enzyme primase.
- Second, DNA polymerases can only move in one direction — from the 3' end of the matrix to its 5' end (in other words, they build up a new DNA strand by lengthening the 3' end of the primer). DNA strands are antiparallel which means that the directions from 3' to 5' on the two strands are opposite. Because of this, one strand replicates continuously (it is called the leading strand) and the other strand replicates in fragments (lagging strand). This will be described in more detail later.

REPLICATION PROCESS

The process begins at the **origin of replication** (also called the replication origin, Fig. 36). The nucleotide sequences of such points differ in different organisms but are usually rich in A and T. Eukaryotic DNA has a large number of origins (bacterial DNA has only one).

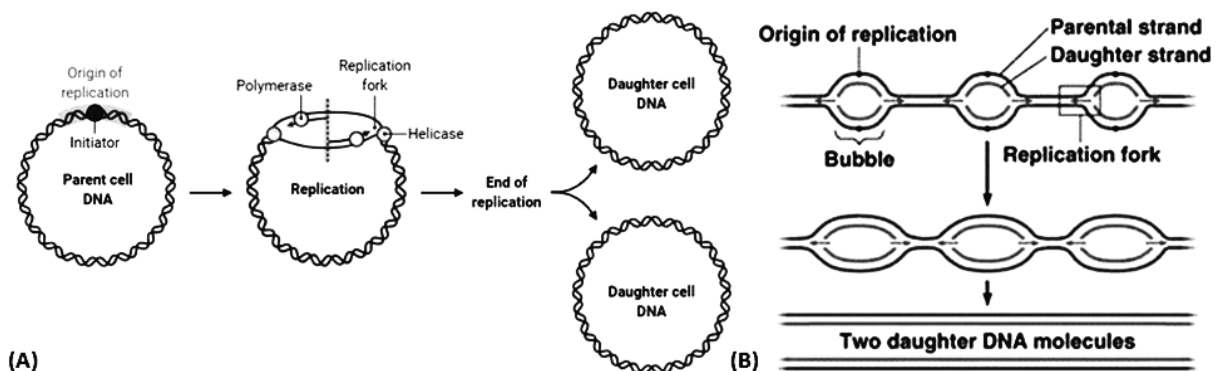


Figure 36. Replication bubbles of prokaryotes (A) and eukaryotes (B)

A DNA fragment that replicates from a single replication origin is called a **replicon**. Since bacteria have a single replication origin in the genomic DNA, their entire circular DNA is one replicon. Eukaryotes have multiple replicons in which replication occurs almost simultaneously, which drastically reduces the time it takes for a DNA molecule to replicate.

For parental DNA strands to serve as templates, they are first separated by **helicase** enzymes. The region of DNA where the strands separate and form a “bloat” called a replication bubble.

The same groups of DNA doubling enzymes move in opposite directions from the origin, so it is enough to illustrate the mechanism of this process on

only one-half of the replication bubble. Such half has a Y shape and is called a **replication fork**.

The replisome moves along DNA, unwinding and separating the strands of maternal DNA. This results in twisting, superspiralization, and tension in DNA, which is eliminated by the enzyme **topoisomerase** (Fig. 37).

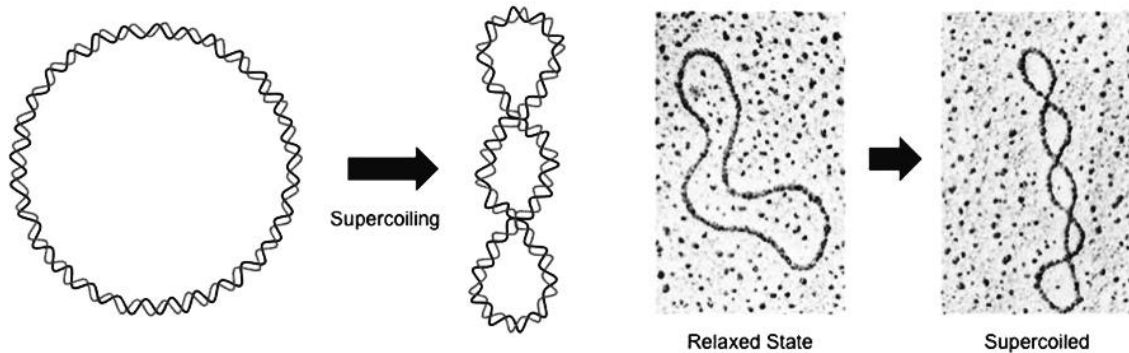


Figure 37. Supercoiling of circular DNA

When helicases unwind maternal DNA at the origin, a specialized RNA polymerase — **primase** — forms a short (~12-nucleotide) complementary to the template strand RNA fragment, called **primer**. Further elongation of the primer by DNA polymerase will be the synthesis of the new daughter strand (Fig. 38).

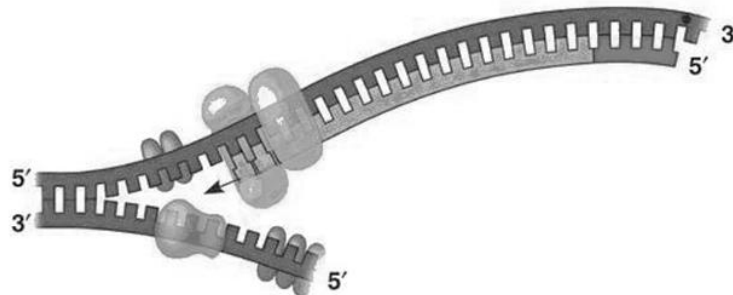


Figure 38. Scheme for the synthesis of a new DNA strand

Due to the antiparallel nature of the DNA strands, when the replisome moves, the unwinding occurs in the 3'→5' direction only for one of the strands. DNA polymerase moves in the same direction (3'→5'). The synthesis of the complementary strand on this template occurs continuously, starting from the primer. This strand is called the **leading strand**. For the second strand (called the **lagging strand**), replisome movement and unwinding occur in the opposite direction 5'→3', so here DNA polymerase, first, moves against the replisome, and second, synthesizes the daughter strand in fragments. Such fragments are called **Okazaki fragments**. A new primer is synthesized on the lagging strand approximately every 100–200 nucleotides (in eukaryotes). Subsequently, the RNA primers are removed and replaced with DNA nucleotides using another type of DNA polymerase. The adjacent Okazaki

fragments are joined by **DNA ligase**. Two enzyme complexes working at the same replication fork are connected in one replisome. They synthesize new chains in opposite directions relative to the template. That's why the lagging strand forms a loop around the replisome in the region of the replication fork. The loop's length is equal to the length of the two Okazaki fragments (Fig. 39).

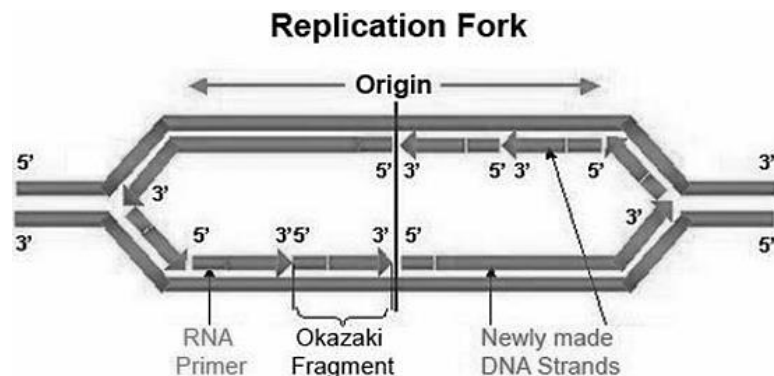


Figure 39. Replication bubble diagram (contains two replication forks separated by a vertical line)

The proteins that perform replication in a single replication fork are assembled into a complex called the **replisome** (Fig. 40).

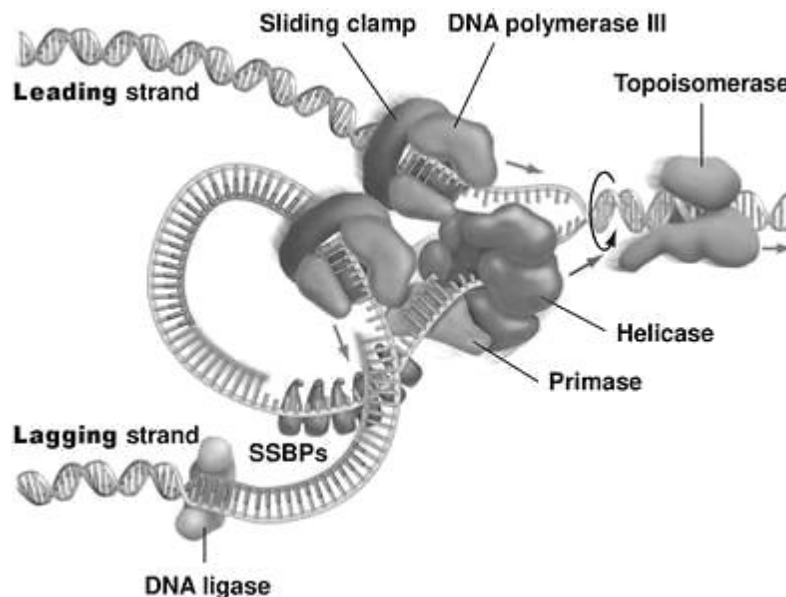


Figure 40. Replisome diagram

Since eukaryotic DNA is linear rather than circular, there is a problem associated with incomplete replication of the DNA ends (telomeres). This is because DNA polymerases cannot start the synthesis of a new nucleotide chain, but only extend the existing fragment, and can move only in one direction, which limits the replication region to the “terminal” primer. The enzyme telomerase is used to elongate telomeres, but it is not active in most somatic cells. This is the reason for the limited number of possible somatic cell divisions.

MITOSIS

Mitosis is an indirect cell division consisting of conventionally defined five phases: prophase, prometaphase, metaphase, anaphase, and telophase (prometaphase is often considered as a part of prophase). At the end of mitosis, the cytoplasmic division between daughter cells occurs — cytokinesis, which is often seen as part of telophase (Fig. 41). During mitosis, mRNA synthesis stops, and protein synthesis decreases and can be about 25 % of the interphase level.

1. The beginning of **prophase** is considered to be the appearance of the first chromosomes distinguishable in the light microscope. In this phase, the process of chromatin condensation, which began in G₂, continues. Centrosomes begin to move in opposite directions, as microtubules begin to proliferate between them. This is how the division spindle, a complex of microtubules, begins to form. The division spindle participates in the distribution of chromosomes between daughter cells in subsequent phases. During the formation of the division spindle, the nuclear envelope is destroyed and the endoplasmic reticulum absorbs its components. At this point, microtubules already stretch through the entire cell from one pole to another.

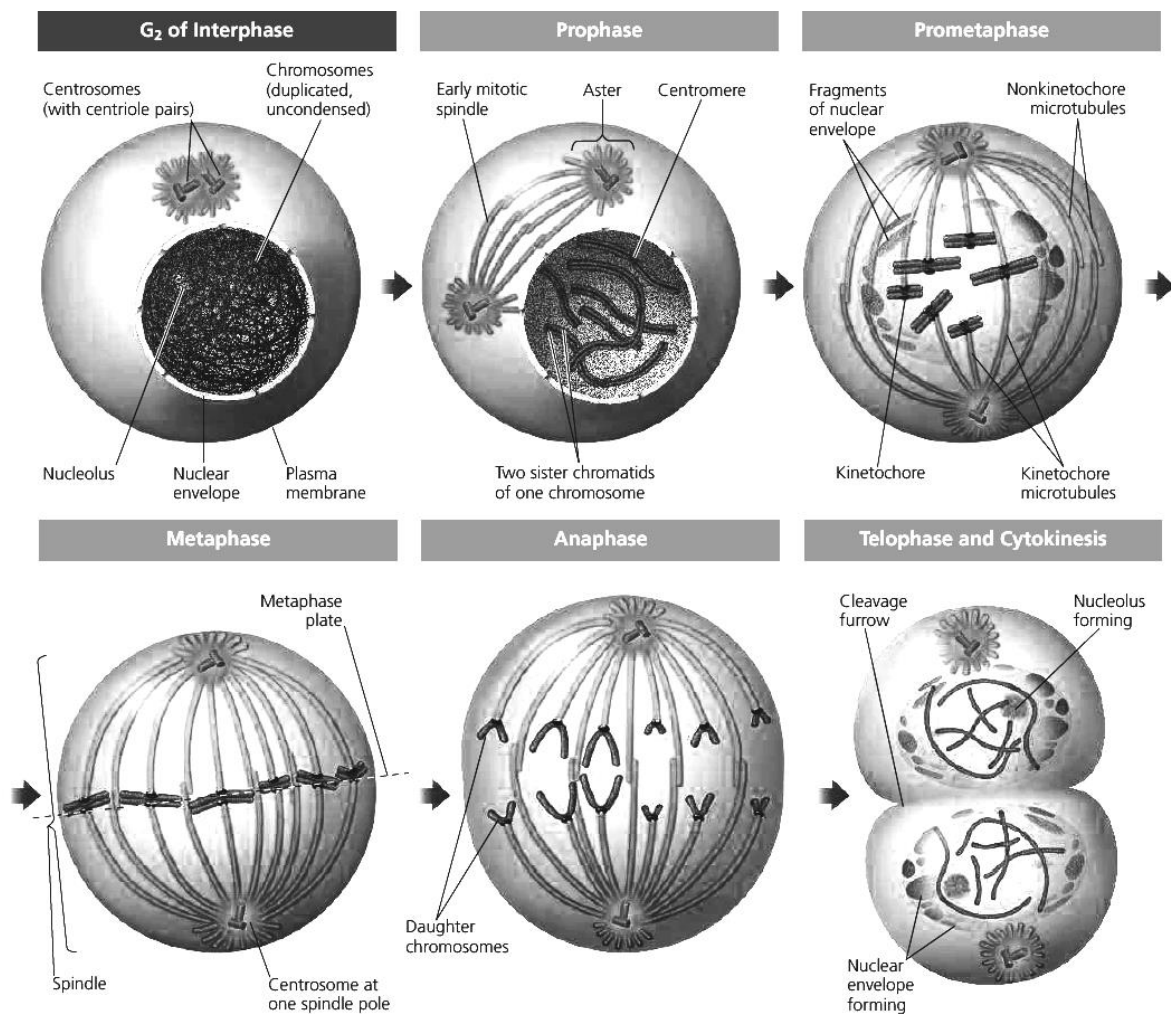


Figure 41. Mitosis in an animal cell

2. **Prometaphase** begins after the fragmentation and disappearance of the nuclear envelope. Condensed chromosomes attach to the microtubules of the division spindle with their kinetochores (protein structures located in the centromere region). Each chromosome has two kinetochores, one at the centromere region of each chromatid. They are attached to microtubules growing from opposite poles. Chromosomes move towards the equator of the cell, tending to occupy an equidistant position from both centrosomes. This process is ensured both by changes in the length of microtubules and by motor proteins located in the region of the kinetochore.

3. **Metaphase** occurs when chromosomes are arranged on the equator of the cell, forming a circle. An imaginary plate passing through this circle and perpendicular to the axis of the division spindle is called the metaphase plate. The sister chromatids of each chromosome are still attached to microtubules extending from opposite poles. With this arrangement, the cell is ready for the sister chromatids to divide properly so that each daughter cell receives a complete set of chromosomes.

4. **Anaphase** is the shortest of all stages of mitosis. By the end of this phase, the content of genetic material at cell poles (i.e., in future daughter cells) is $2n1chr2c$. Previously, it was thought that in anaphase there is a rapid replication of centromere regions under-replicated in interphase. It has now been established that before anaphase, sister chromatids are held together by cohesin proteins concentrated near centromeres. The key event of anaphase is the simultaneous removal of these proteins, which is considered to be its beginning. Having separated from each other, sister chromatids quickly pull up to the poles to which their kinetochores are attached by microtubules.

5. In **telophase**, the division spindle is disassembled into tubulin monomers, which can be used to build the cytoskeletons of daughter cells. Around each set of daughter chromosomes (former sister chromatids of the mother cell), nuclear envelopes are formed. The chromosomes become decondensed, lose their clear outlines, and form the chromatin of the interphase nucleus. The expression of genes, including rRNA genes, begins, leading to the appearance of the nucleolus. Telophase can be considered as a process opposite to prophase, which returns the cell to interphase.

6. **Cytokinesis** is the division of the cytoplasm of the mother cell between daughter cells. In animal cells, cytokinesis occurs using a contractile ring, which is usually formed in the equatorial plate of the division spindle. This ring contains actin microfilaments as well as myosin. When contracted, the contractile ring forms a constriction that divides the cell in two. After cytokinesis, the cell enters the G1 phase.

Mitosis helps to ensure that both daughter cells contain a complete set of chromosomes, but does not guarantee that organelles such as mitochondria and chloroplasts are also evenly distributed. However, if at least one of these organelles is present in each cell, it can replicate until it reaches the required number.

AMITOSIS

Amitosis is a form of eukaryotic cell division by direct division of the nucleus and division of the cytoplasm. There is no condensation of chromatin into chromosomes or fragmentation of the nuclear envelope, and the division spindle is not formed. It is also assumed that there is no uniform distribution of genetic material between the daughter nuclei. This phenomenon was discovered in the 19th century, and the term “amitosis” was proposed by Walter Flemming. Amitosis is much less well-studied than mitosis. According to the available data, in eukaryotes, amitosis occurs mainly in polyploid cells, which are common in many unicellular cells (ciliated protists), plants, as well as in some tissues of diploid animals, including humans.

Endoreduplication. In many organs and tissues of normal diploid animal and plant organisms, polyploid cells are found. The process of formation of such cells is called endoreduplication. An increase in the ploidy of a cell can occur in a variety of ways, among them:

- **Endomitosis.** At the beginning of endomitosis, chromosomes condense and become clearly visible within the nucleus. Then, visible chromosomes disappear, and the nucleus takes the form of a normal interphase nucleus, but its size increases in proportion to the content of genetic material.

This is how polyploid cells can be formed in many plants, protists, and animals.

- **Polytenia.** In the S phase, replication occurs, but the daughter chromosomes remain located near each other in a decondensed state, then the next replication cycle occurs, without divergence, and so on. Replication results in the formation of polytene chromosomes. They are several times thicker than mitotic chromosomes, as they are bundles of multiple unexpanded chromatids, and 70–250 times longer than mitotic chromosomes, as they are weakly condensed. Polytene chromosomes can be found in the salivary gland cells of many insects. The biological meaning of this phenomenon may be related to an increase in the number of gene copies, which also increases the number of products of these genes.

REGULATION OF THE CELL CYCLE

Preparation for mitosis or transition to the G₀ phase is regulated by various stimuli, both external to the cell and internal. The ability to divide is affected by signals from other cells, such as growth factors. Growth factors are a large group of molecules (most often proteins) secreted by different cell types and stimulate division. They bind to certain cellular receptors and cause a chain of interactions between signaling molecules within the cell, which can ultimately induce mitosis. Other examples of factors regulating division are the availability of nutrients (if they are lacking, the division may not occur),

free space (contact with enough neighboring cells prevents division), connection to the extracellular matrix (many cells do not divide if not attached to the substrate) and other factors. There are experiments that a decrease in the nuclear-cytoplasmic ratio is also associated with the beginning of division.

Checkpoints of the cell cycle. In the cell cycle, there are certain stages in which the possibility of the transition to the next phase is actively controlled. Such stages are called cell cycle checkpoints. Internal and external conditions are evaluated there, and if they are unfavorable for cell division, the cell is retained in the current phase. The most important checkpoints are G1, G2, and M (the checkpoint of the division spindle assembly).

1. **The G1 (or G1/S) checkpoint** is the primary point where the cell “determines” whether to divide or not. Passage of this checkpoint and transition to the S-phase is greatly facilitated by growth factors — signaling molecules that bind to cell receptors and activate certain signaling pathways in the cell, stimulating preparation for mitosis (e.g., platelet-derived growth factor released during injury from activated platelets). The presence of sufficient nutrients and the absence of DNA damage are also important conditions.

2. **The G2 checkpoint (G2/M)** occurs at the end of phase G2. Here the completeness of DNA doubling is assessed because mitosis cannot be started until replication is complete. In addition, there is a check for DNA damage, and if it is found, the cell also does not proceed to mitosis until repair (i.e., DNA repair — see Chapter 13) is complete.

3. **The M checkpoint** (the checkpoint of the division spindle assembly) occurs at the end of the metaphase. For its successful passage, all chromosomes must be attached to the microtubules of the division spindle, because if this condition is not met, the division of chromosomes between daughter cells can occur in error.

THE ROLE OF CYCLINS AND CYCLIN-DEPENDENT KINASES IN CELL CYCLE REGULATION

Recall that various events occurring in the cell usually involve proteins. A common way to activate or inactivate the functions of various proteins is the addition of phosphoric acid residues (phosphorylation). Proteins that attach phosphate groups to other molecules are called **kinases**. Enzymes that regulate the work of proteins that perform certain functions in different phases of the cell cycle are called cyclin-dependent kinases (Cdk).

These proteins are called Cdk because their activity is regulated by cyclins. **Cyclins** are other proteins whose concentration changes cyclically depending on the phase of the cell cycle (the concentration of Cdk does not change in the same way). There are different cyclins (D, E, A, B) and different **cyclin-dependent kinases** (Cdk1, Cdk2, Cdk3, etc.). When the concentration

of certain cyclin increases, it binds to a certain type of kinase (or several types). The resulting complexes phosphorylate certain target proteins, ultimately leading the cell to perform the actions required during a given phase of the cell cycle. To illustrate this mechanism, let us look at an example. Growth factors interacting with the cell in the G1 phase can act as signals for the synthesis of cyclin D, which binds to Cdk4 and Cdk6. The resulting complexes phosphorylate (and thus “turn off”) protein Rb (retinoblastoma protein), which releases E2F transcription factors (transcription factors regulate other genes, their work will be discussed in detail in Chapter 6), which are important for the synthesis of proteins that provide the transition to the S phase. Among these proteins are other cyclins.

Disorders of cell cycle control in a cell can lead to uncontrolled cell division and tumor formation. This will be discussed in more detail in Chapter 13.

APOPTOSIS AND NECROSIS

Cell death can occur in different ways, the most studied of which are necrosis and apoptosis.

Necrosis can be thought of as uncontrolled cell death. It occurs when cells are exposed to injury or excessive stresses, such as high temperature, lack of oxygen, or microbial contamination. Because of damage to the plasma membrane, the cell loses its ability to control ion and water flow, swells, cytoplasmic acidification occurs, vacuolar components swell and rupture their membranes, protein synthesis in the cytosol is stopped, lysosomal hydrolases are released and lysis of the cell occurs. Autolysis, the process of cell self-digestion under the action of its enzymes, is observed. The peculiarity of necrosis is that large groups of cells undergo such death (for example, in myocardial infarction (heart attack) due to lack of oxygen supply to a part of the heart muscle). As a rule, cells that undergo this process swell and burst, releasing their intracellular contents, which can damage the surrounding cells and often cause inflammation.

Apoptosis is programmed and controlled cell death. In contrast to necrosis, apoptosis does not chaotically destroy the cell, and the cell contents are not released to the outside, adjacent cells are not damaged, and there is no inflammation. Instead, the cell disintegrates into vesicles called apoptotic bodies, after which they are phagocytized by macrophages or adjacent cells. Apoptosis is observed during the normal development of the organism to remove unnecessary embryonic structures such as Wolffian and Müller canals, the tissue between the fingers of the embryo, metamorphosis of insects and amphibians. Mutant or infected cells are “self-destructed” by apoptosis. In the immune system, apoptosis removes those maturing lymphocytes that can attack the body’s normal cells, and cells attacked by lymphocytes (such as

cancer cells or infected by viruses) also die by apoptosis. The key proteins of apoptosis are **caspases**. These proteins belong to proteases (enzymes that break down proteins) that break peptide bonds after asparagine acid. They are constantly present in the cell in an inactive state and are activated when needed. A large number of different cellular proteins are broken down by caspases, eventually leading to the destruction of the cytoskeleton, the nuclear lamina, and other substrates. Chromatin condenses, the nucleus fragments, a distinctive “bubbling” of the membrane is observed, and the cell disintegrates into apoptotic bodies.

There are different pathways of apoptosis activation. Two of them are the most well-characterized: intrinsic (mitochondrial) pathway, and extrinsic (receptor-dependent), activated from outside. The mitochondrial pathway is associated with the release of apoptosis-related proteins from mitochondria, which eventually leads to caspase activation. This is how apoptosis is most often realized in animal cells. The extrinsic pathway is induced by the binding of certain single molecules to cell death receptors.

BINARY FISSION IN BACTERIA

Despite the differences between prokaryotic and eukaryotic cells, the essence of their division process is the same: the genetic material doubles and, together with the cellular content, divides between the daughter cells. Due to the simpler structure of a prokaryotic cell, its division mechanism is also simpler.

In some bacteria, the division is initiated when the circular DNA — nucleoid — begins to double at the origin of replication. When replication is complete and the bacterium reaches about twice its original size, the parent cell divides into two daughter cells. At the time the chromosome replicates, the cell elongates. Previously, it was thought that newly replicated DNA molecules of bacteria such as *E. coli* were separated passively due to cell growth and distancing of the membrane regions to which they were attached. However, it has been found that after DNA doubling, the replication origins are removed from each other at a higher speed than cell growth occurs, indicating the presence of mechanisms for the separation of genetic material.

MEIOSIS

Animal development begins with the division of a diploid zygote cell, which is formed by the fusion of two parental cells — gametes. Gametes are haploid cells ($1n$), i.e., they contain only half the set of chromosomes typical of somatic cells of the species. This prevents the chromosome set from doubling with each generation. Because of this, animals have a way of producing haploid cells from diploid ones. This process is meiosis.

The mechanism of meiosis is in many ways identical to that of mitosis. The key differences between these processes are:

- Mitosis is a single division in which sister chromatids are separated between two daughter cells. Meiosis consists of two divisions: during the first one, homologous chromosomes rather than chromatids are separated between the two cells; the second division is similar to mitosis.

- In the prophase of the first division of meiosis, homologous chromosomes are joined into a single structure (bivalent), just as sister chromatids are joined during mitosis. This is necessary for the unmistakable separation of each pair's chromosomes between daughter cells.

- The biological meaning of mitosis is to increase the number of cells; the primary biological meaning of meiosis is the formation of haploid cells from diploid cells.

Stages of meiosis. Meiosis includes two divisions, meiosis I (reduction division) and meiosis II (equational division). The short period between them is called interkinesis. Each division consists of prophase, metaphase, anaphase, and telophase.

Starting meiosis, the cell contains a set of chromosomes $2n2chr4c$, i.e., it has a double set of chromosomes, each consisting of two sister chromatids. Recall that sister chromatids contain identical DNA — copies of the same molecule formed as a result of replication. Homologous chromosomes contain non-identical DNA molecules — they have the same genes at the same loci, but the alleles of these genes may be unequal (alleles are different versions of the same gene). For example, a particular locus of the ninth chromosome of one parent may contain the gene responsible for the development of the second blood group, while the same locus of the ninth chromosome of the second parent contains the third.

1. Prophase I is the most complicated stage. As in mitosis, the centrosomes move to the cell poles, the division spindle is formed, and the nuclear envelope is destroyed. The sister chromatids of each chromosome are held together by cohesin proteins. However, unlike mitosis, synapsis and crossing over occur during this period.

- **Synapsis** is the gradual joining of two homologous chromosomes along their entire length. Each gene on one chromosome is placed exactly opposite the same gene on the other. The homologous chromosomes are held together by a protein complex called the synaptonemal complex (SC). A pair of homologous chromosomes connected in this way is called a *bivalent* or *tetrad*. The content of genetic material in a cell can be written not only as $2n2chr4c$, but also as $1n_{biv}4chr4c$, which means that the cell contains $1n$ bivalents, and each of them has four chromatids, the cell has 4 sets of DNA.

- **Crossing over** is the process by which homologous chromosomes exchange identical regions. The DNA of two adjacent, but not sister chromatids

(i.e., similar maternal and paternal chromatids) is broken by specific proteins at the same points. Further, the DNA breaks are closed in such a way that each end of the DNA fragment is attached to an adjacent segment of a non-sister chromatid. Thus, parts of the paternal chromatid are joined to parts of the maternal chromatid and vice versa.

Then, after the synaptonemal complex is disassembled, the homologous chromosomes slightly diverge, and the chromatid junction points begin to look like crossings of homologous chromosomes. Such crossings are called chiasmata. Homologous chromosomes remain attached to each other due to cohesins, even though some DNA may no longer be attached to “their” chromosome. There must be at least one crossing over between chromosomes for a pair to stay together during the transition to metaphase I, but there are known exceptions. For example, in male *Drosophila* the crossing-over does not occur, but meiosis proceeds in a normal way. The mechanism holding the homologous chromosomes together, in this case, is not fully understood.

Due to the complexity of prophase I, it is divided into five stages (Fig. 42): **leptotene** (visible chromosome condensation begins), **zygotene** (synapsis begins), **pachytene** (crossing over occurs), **diplotene** (crossing over is completed, the synaptonemal complex is removed, chiasmata are visible) and **diakinesis**.

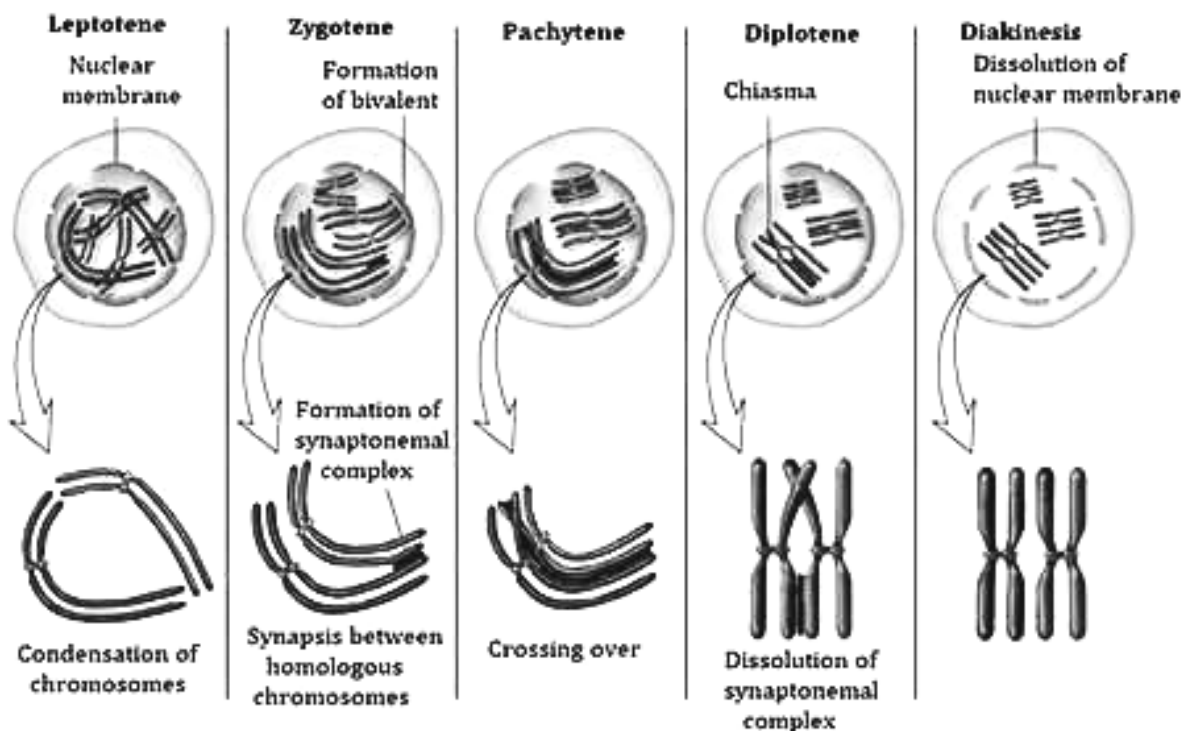


Figure 42. Stages of prophase I of meiosis

2. Metaphase I ($1n_{biv}4chr4c$). With the help of the division spindle microtubules, bivalents move to the equator of the cell, forming the metaphase plate. In this case, microtubules growing from opposite poles attach to

homologous chromosomes rather than to sister chromatids, as occurs during mitosis. The orientation of each bivalent is random — both maternal and paternal chromosomes can be oriented to any pole of the cell.

3. **Anaphase I**. Microtubules begin to shorten, pulling homologous chromosomes toward opposite poles. Anaphase I occurs by removing cohesins from the arms of the chromosomes, but not from the centromeres. As a result, each pole of the cell contains chromosome set $1n2chr2c$, which means that the cell contains $1n$ chromosome, and each of them has 2 chromatids, the cell has 2 sets of DNA.

4. **Telophase I** ($1n2chr2c$). The nuclear membrane appears and daughter nuclei are formed. Cytokinesis, the division of the cytoplasm and its contents, may or may not occur after telophase I. The next division occurs after interkinesis.

Meiosis II proceeds similarly to mitosis.

5. **Prophase II** ($1n2chr2c$). Each nuclear envelope is destroyed and the division spindle is formed.

6. **Metaphase II** ($1n2chr2c$). Chromosomes are lined up at the cell equator.

7. **Anaphase II** ($1n1chr1c$). Microtubules are shortened, and cohesins connecting centromeres of sister chromatids are removed, allowing sister chromatids to diverge toward opposite poles, as in mitosis.

8. **Telophase II** ($1n1chr1c$). The nuclear envelope forms around four sets of daughter chromosomes, and cytokinesis occurs. The result of this division is four haploid cells. Each of them can develop directly into gametes, as occurs in animals, or divide mitotically, as occurs in plants, fungi, and many protists.

Meiosis I is much more prone to errors compared to mitosis. Any failure during meiosis, such as a chromosome mismatch, results in gametes with an abnormal chromosomal set. Such gametes are called aneuploid gametes. In humans, this condition is the most common cause of spontaneous abortions. According to some estimates, about 10 % of all human pregnancies involve aneuploid gametes.

MEDICAL ASPECTS OF CELL PROLIFERATION

Cell proliferation is the increase in the number of cells as a result of cell growth and division. The process of cell proliferation is the basis of both vital and pathological processes in the body. Understanding the mechanisms of cell proliferation is necessary for:

- Prevention of birth defects.
- Methods of maintaining body functions, organ and tissue repair after damage resulting from diseases, injuries, exposure to toxins, etc.
- Development of methods to increase longevity.
- Prevention and treatment of cancer.

CHAPTER 5

THE FLOW OF GENETIC INFORMATION IN THE CELL

PROPERTIES AND FUNCTIONS OF GENES

Gene is an association of genomic sequences encoding a coherent set of potentially overlapping functional products as part of genetic regulatory networks.

The primary functions of genes are the storage and transmission of genetic information.

Genes are characterized by the following properties:

- Gene specificity means that each gene determines the synthesis of a strictly defined protein (or RNA).

- Gene integrity is that gene acts as an indivisible unit when programming polypeptide synthesis. The gene as a functional unit is indivisible.

- The discreteness of a gene is determined by the presence of subunits in it. A pair of complementary nucleotides can be considered the smallest structural unit of a gene, and a codon can be considered the smallest functional unit.

- The stability of a gene is that the information encoded in it usually does not change. Nevertheless, with some frequency, such changes (mutations) can occur. The ability of a gene to change (mutate) is called lability.

- Genes usually have a pleiotropic (multiple) effect, i.e., one gene is responsible for the expression of several traits. This phenomenon, in particular, is observed in some hereditary diseases, especially those associated with connective tissue pathologies.

- The phenotypic expression of a gene can be influenced by environmental factors and other genes. The degree of phenotypic expression of a gene is called expressivity. The degree and expressivity of genes are important in medical genetics. The varying severity of the disease course in different individuals is explained by this property of the gene.

- The frequency of gene expression is called penetrance. It is expressed as a percentage of the number of individuals who have the trait to the number of individuals who have the gene.

THE STRUCTURE OF EUKARYOTIC GENES

Unlike bacterial genes, eukaryotic genes usually have an “interrupted” structure — consisting of protein-coding sections of exons and introns separating them and not coding for protein.

In humans, the sum of all exons of protein-coding genes is about 1.5 % of the genome, while the sum of introns is about 24 %. The human gene contains on average about 9 exons and 8 introns. In humans, the average intron length is 3.3 kb, but most are less than 200 bp long.

At first glance, the presence of introns in eukaryotic genes may seem meaningless, but in reality, it is not. Among the functions of introns are the following:

- Separation of exons opens up the possibility for alternative splicing, a way of increasing the protein diversity of an organism, described further below.
- The presence of introns opens the evolutionary pathway for the appearance of new genes by combining exons during the mutation process.
- Introns are involved in the regulation of gene expression in various ways. For example, they can contain alternative promoters, and enhancers (described in the next chapter).
- Introns may contain sequences of noncoding RNAs, which can include microRNAs, small nuclear RNAs, and small interfering RNAs.
- Because transcription of a large intron takes time (often several minutes), they can provide a delay between gene activation and the appearance of its protein product in the organism. This regulatory mechanism is often found in genes involved in organism development.

THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology is the concept that explains how information stored in nucleic acids is realized in the cell. It was first proposed by Francis Crick in 1957. Its point was that based on DNA, RNA is created, and based on RNA, protein is created ($\text{DNA} \rightarrow \text{RNA} \rightarrow \text{protein}$). At present, the central dogma of molecular biology can be depicted as follows: $\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{protein}$. Let us consider each direction of transmission of hereditary information in more detail:

- Information can be copied from the maternal DNA to the daughter (DNA). This process is called **replication**. The key enzyme performing replication is **DNA polymerase**. This process takes place in all living organisms.
- Information can be copied from DNA to RNA ($\text{DNA} \rightarrow \text{RNA}$), this process is called **transcription** and is performed by **RNA polymerases**. Transcription is a necessary step in protein synthesis, so it occurs in all living organisms. However, not all types of RNA are used in this role, as they perform other functions.
- Information can be copied from RNA to DNA ($\text{RNA} \leftarrow \text{DNA}$); this process is called **reverse transcription** and is performed by **reverse transcriptases (revertases)**. This pathway is not widespread but is used by some viruses and mobile genetic elements (discussed in the next chapter).
- Some viruses can copy information from the mother RNA to the daughter RNA ($\text{RNA} \rightarrow \text{RNA}$). This process is called **RNA replication** and is performed by **RNA replicases**.

- The information recorded in RNA can be used for protein synthesis (RNA → protein). This process is called **translation** and is performed by the **ribosome**.

- Information transfer directly from DNA to protein as well as from protein to RNA or DNA is not found in nature.

The mechanisms described below will be described in more detail.

THE STRUCTURE AND FUNCTION OF RNA

RNA (ribonucleic acid) has a similar structure to DNA. However, there are significant differences:

- RNA nucleotides have the five-carbon sugar ribose rather than deoxyribose (the other components of the nucleotides are the same — phosphate and nitrogenous base). This makes RNA less stable than DNA (spontaneous hydrolysis of the phosphodiester bond between nucleotides is possible).

- RNA nucleotides have **uracil** (U) instead of thymine (T). These nitrogenous bases have a similar structure and differ only in the presence of one methyl group in thymine. Uracil can also form complementary pairs with adenine. The remaining bases (A, G, C) are present in both DNA and RNA.

- RNA (except in some cases) is a single strand. However, complementary pairs can form in it, causing RNA to fold and acquire a secondary structure.

- RNA that has a secondary structure can have enzymatic activity, which is not typical of DNA. Such “RNA enzymes” are called **ribozymes**.

- The size of DNA usually exceeds the RNA size. RNA of various types contains several hundred or thousands or often only a few dozens of bases. The function of DNA is to store hereditary information in the cell as it functions and as it is passed on from generation to generation. Since RNA is composed of nucleotides, it can also perform a similar function (e.g., many viruses), but most types of RNA perform other functions.

There are different types of RNA, most of which are noncoding (i.e., do not carry information about protein structure):

1. **Messenger RNA (mRNA)**. This type of RNA carries information about protein structure from DNA to the ribosome.

2. **Ribosomal RNA (rRNA)** — several types of RNA that are key components of ribosomes. The formation of peptide bonds between amino acids is catalyzed by one of the rRNAs.

3. **Transfer RNA (tRNA)** is a type of RNA that delivers amino acids to the ribosome. Each type of tRNA can be associated with only one type of amino acid. When it enters the ribosome, the tRNA is checked for “correctness” and the amino acid is attached to the growing polypeptide chain. Thus, the tRNA is an “adapter” between the mRNA and the amino acid.

4. **Small nuclear RNA (snRNA)**. These RNAs are part of the mechanism involved in the processing of immature mRNAs (**pre-mRNAs**).

5. **MicroRNAs (miRNAs)** and **small interfering RNAs (siRNAs)** are involved in controlling gene expression.

There are also a large number of RNA types beyond those described. The functions of the various RNAs will be described further in this and subsequent chapters.

GENETIC CODE

The recording of information in the nucleotide chain is possible since it is a sequence of four types of nucleotides (A, G, C, T/U). One amino acid in this system is “written” by three nucleotides (e.g., AUG is methionine, CCC is proline, etc.). The triplet of adjacent nucleotides encoding an amino acid is called a codon. So, the sequence of codons in a gene determines the sequence of amino acids in the protein encoded by this gene. The system of recording information about the protein structure by the nucleotide sequence of DNA or RNA is called the genetic code (Fig. 43).

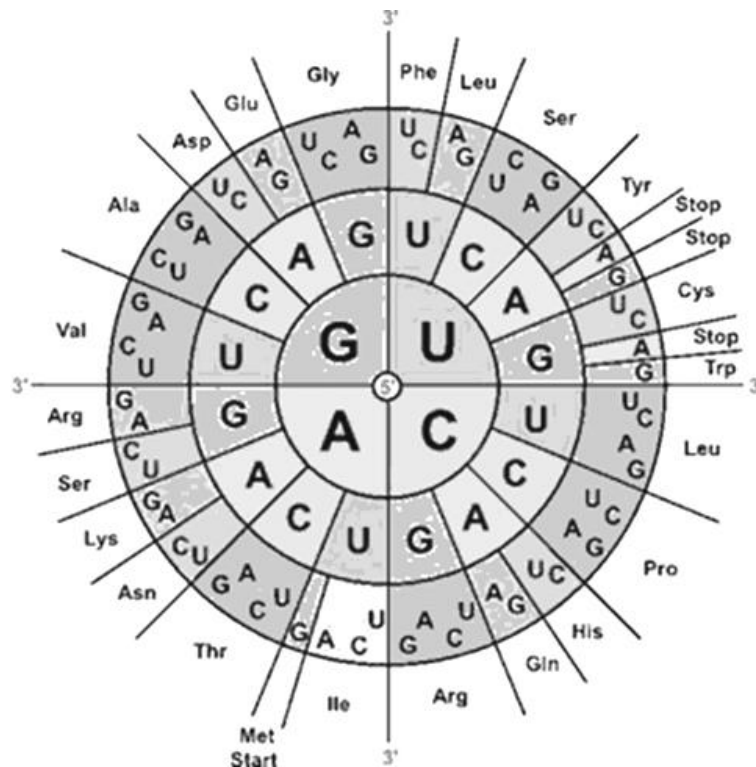


Figure 43. Genetic code

Properties of the genetic code:

- Triplettness — one amino acid is encoded by a triplet (codon) — three adjacent nucleotides.
- **Redundancy, or degeneracy.** Since the total number of all possible codons is $4^3 = 64$ the number of amino acid types is 20, the same amino acid can be encoded by several different codons. For example, both the mRNA UUU and UUC codons encode phenylalanine.

- The existence of a start codon and stop codons. The start codon in mRNA is the triplet from which translation begins: AUG (it also encodes methionine). The stop codons (UAA, UGA, UAG) mark the endpoint of translation. They do not encode any amino acids, so they are also called nonsense codons.

- There are no “separation marks” between codons, so removing or adding a nucleotide to the coding sequence will lead to the reading frameshift.

- Non-overlapping — one nucleotide cannot be a part of several adjacent codons at the same time.

- Single direction — DNA and RNA polymerases in replication and transcription always move from the 3' to the 5' end of the template strand; the ribosome moves from the 5' to the 3' end of mRNA.

- **Universality** — the genetic code is the same in all living things (with some minor exceptions in some organisms).

TRANSCRIPTION

Transcription is the process of creating an RNA copy of a DNA region. It is performed by RNA polymerase and is based on the principle of complementarity. During this process, two DNA strands are separated at a certain site, and one of them serves as the matrix for RNA synthesis. The complete set of all RNA transcripts, including non-coding ones, characteristic of a given organism, or a specific set of transcripts (RNA molecules) present in cells of a certain type is called transcriptome. It will be described in more detail in the next chapter.

Transcription in prokaryotes. In bacteria, transcription occurs in the cytoplasm. The resulting transcript (i.e., mRNA) does not require modifications, and translation begins before transcription is complete.

Transcription consists of three steps:

- **Initiation.** At this stage, the RNA polymerase binds to a specific region of DNA upstream of the region to be transcribed (Fig. 44). This region of DNA, which marks the start region of transcription, is called the promoter. RNA polymerase temporarily separates DNA strands in a small area, forming a **transcription bubble**.

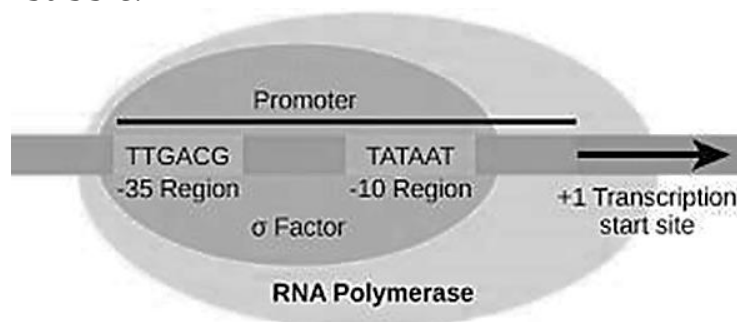


Figure 44. Diagram of RNA polymerase located on the promoter at the initiation of transcription

- **Elongation.** At this stage, the RNA is directly assembled on the DNA matrix. DNA has two strands. The mRNA has the same nucleotide sequence as one of them but is complementary to the other. The DNA strand that has the same sequence as the mRNA is called a **sense strand**, or **coding strand**. However, it is assembled on the other, complementary strand. Because of its complementarity to mRNA, this strand is called **antisense** (or **noncoding**). During elongation, the RNA polymerase moves from the promoter in the 3' → 5' direction relative to the antisense strand. As it elongates, the synthesized RNA (transcript) is separated from the DNA strand and exits the RNA polymerase through a special channel. Only a small number of nucleotides at its 3' end are permanently complementary to the DNA (Fig. 45).

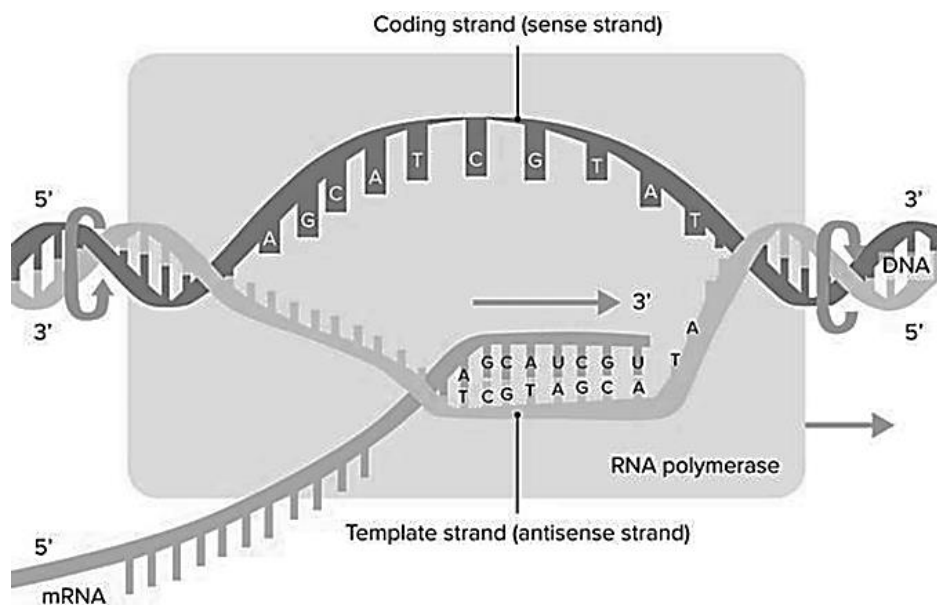


Figure 45. Schematic representation of the elongation process during transcription

- **Termination**, i.e., the end of transcription occurs when the RNA polymerase reaches a DNA region called the terminator. As a result, the formed RNA is released, and the hydrogen bonds between the DNA strands are restored (Fig. 46).

Transcription in eukaryotes and pre-mRNA processing. Eukaryotes have several types of RNA polymerases producing different types of RNA, while bacteria have only one. The promoters to which they bind are different. All mRNAs are transcribed by RNA polymerase II.

Let's look at the work of RNA polymerase II, which synthesizes mRNA, the first step in protein synthesis.

In eukaryotes, transcription occurs in the nucleus, and therefore translation occurs after the transport of mRNA into the cytoplasm. The transcript formed by RNA polymerase is not yet mature mRNA and is called **pre-mRNA**. Before translation, pre-mRNA must undergo **processing**. Processing includes **capping**, **polyadenylation**, and **splicing**, described below.

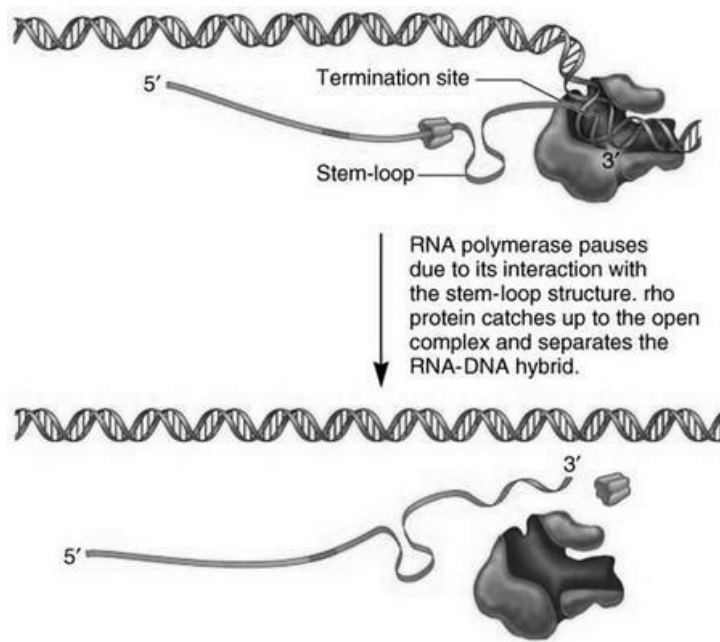


Figure 46. One of the variants of transcription termination in prokaryotes

Initiation on RNA polymerase II promoters is similar to that in prokaryotes, but eukaryotes use multiple **transcription factors** instead of a single protein factor that provides promoter recognition. Transcription factors are proteins that help to attract RNA polymerase II to the promoter by their action. They interact with RNA polymerase II to form the **preinitiation complex (PIC)** on the promoter (Fig. 47). This process will be discussed in more detail in the next chapter.

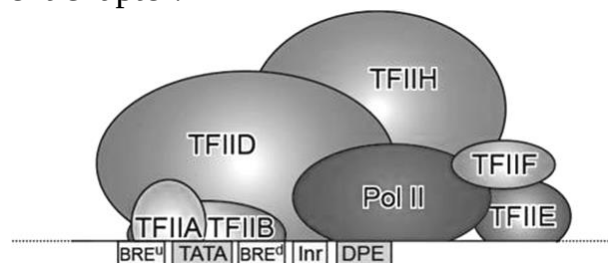


Figure 47. Initiation of transcription in eukaryotes

When the transcript reaches the length of about 20 nucleotides, the formed fragment is **capped**: GTP is attached to its 5' end and then methylated. This results in the formation of a **5'-cap** consisting of triphosphomethylguanosine attached to the transcript with its 5'-end (Fig. 48). This structure is important for translation, RNA stability, and further processing.

Elongation also involves protein factors (e.g., those involved in passing through nucleosome-associated sites). A single gene can be transcribed simultaneously by several consecutive RNA polymerase molecules. This increases the amount of mRNA produced, which helps the cell to produce protein in large quantities.

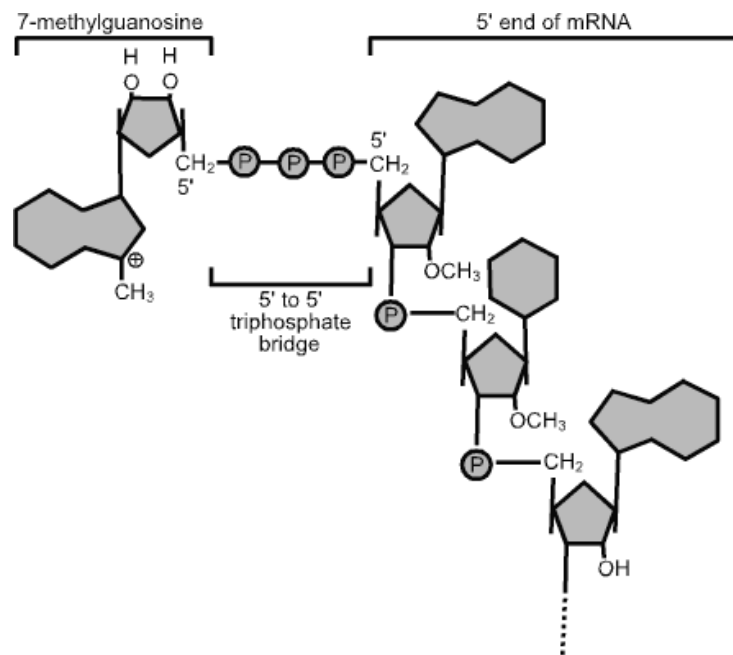


Figure 48. 5'-cap scheme for eukaryotic RNA

The **termination** of transcription in eukaryotes differs from that in bacteria. RNA polymerase II transcribes the polyadenylation signal sequence. Next, specific proteins cut off the transcript from the RNA polymerase, releasing the pre-mRNA. **Polyadenylation** then occurs — 100–200 adenosine residues are added to the resulting 3' end of the transcript, forming a **3' poly-A tail**. The enzyme responsible for this is called poly-A polymerase.

Next, splicing is completed and RNA is transported into the cytoplasm.

Cap and poly-A tail have several important functions in common: they facilitate the export of mRNA from the nucleus, protect it from degradation by hydrolytic enzymes, and help ribosomes to attach to the 5' end of mRNA.

SPLICING

Eukaryotic genes consist of **exons** and **introns**. However, RNA is transcribed continuously and therefore contains sequences of both introns and exons. Because of this, it cannot be considered a finished mRNA. Such a molecule is called “**pre-mRNA**” or “**primary transcript**”. After transcription, pre-mRNA undergoes splicing, a process in which introns are removed and exons are joined together.

Splicing occurs in the nucleus and can start before transcription is complete as a transcript of sufficient length is synthesized. It is performed by **spliceosomes**. Spliceosomes consist of RNA and protein and are therefore ribonucleoprotein complexes. First, the boundaries between introns and exons are recognized by small nuclear ribonucleoproteins (snRNPs), complexes containing small nuclear RNA (snRNA), and proteins. The snRNPs then connect to the major protein subunits that form the **spliceosome**. Both ends of

the intron are brought closer together, the intron is excised and closed into a “loop” which is then removed (Fig. 49).

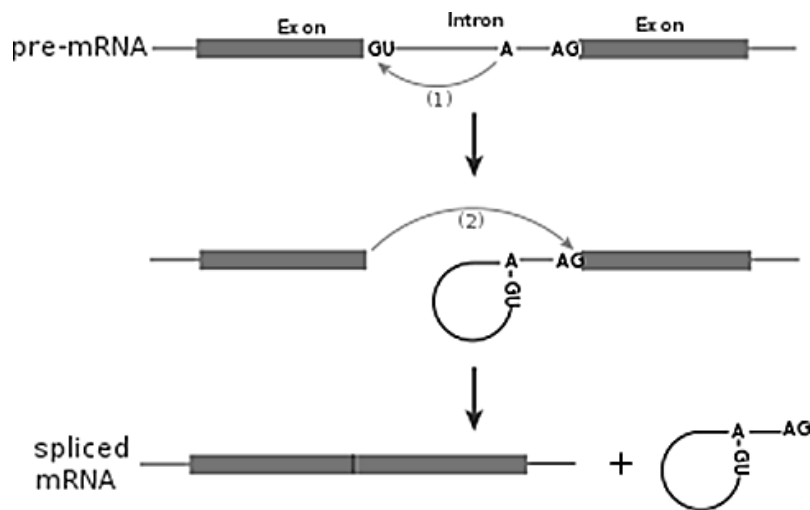


Figure 49. Removal of the intron from pre-mRNA during splicing

Alternative splicing. In eukaryotes, more than one mRNA variant can be produced from the same pre-mRNA through a process called **alternative splicing** (Fig. 50). Therefore, a single eukaryotic gene can encode several protein variants at once. Such variants are called **isoforms**. Because of this, the protein diversity of the organism can be high, while the size of the genome will not increase significantly. In humans, about 95 % of genes containing multiple exons undergo alternative splicing.

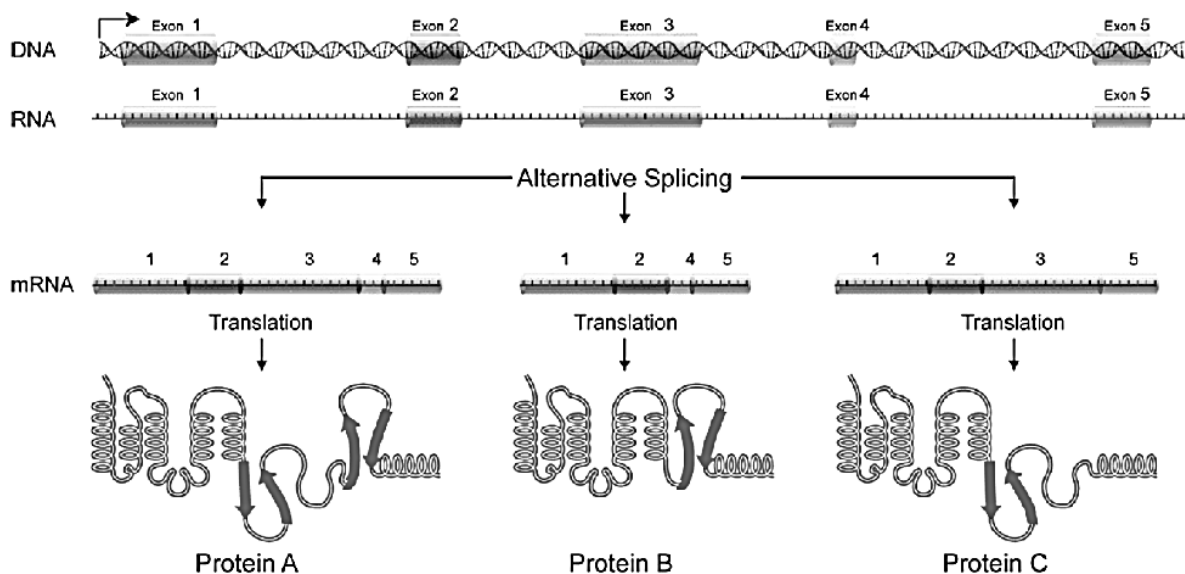


Figure 50. Alternative splicing

Alternative splicing can occur in two main ways:

- Removal of some exons along with introns.
- A gene can have several mutually exclusive exons, and only one of them always gets into the mRNA.

Many inherited diseases can be caused by splicing mutations. These include thalassemias, muscular dystrophies, Hutchinson-Gilford progeria, and a large number of other diseases. The next step in the synthesis is mRNA transport to the cytoplasm, where translation occurs.

TRANSLATION

The **translation** is the process of assembling a polypeptide (protein) from amino acids by the ribosome according to the nucleotide sequence in the mRNA. The entire set of proteins expressed in a given cell type or an organism, at a given period under given conditions, is called the proteome. It is discussed in more detail in the next chapter. Translation requires:

- Amino acids are the structural components of a protein. Recall that a protein is a chain of amino acids (polypeptide) arranged in a specific three-dimensional structure capable of performing specific functions. This structure is determined by the varieties of amino acids that make up the protein and their arrangement in the chain.
- Messenger RNA (mRNA) contains a nucleotide sequence that encodes the amino acid sequence of a protein. In a sense, mRNA can be considered the “instruction” for assembling a protein. The start and stop codons in mRNA are located at a certain distance from the 5' and 3' ends, so each such molecule has non-translatable end regions (Fig. 51).

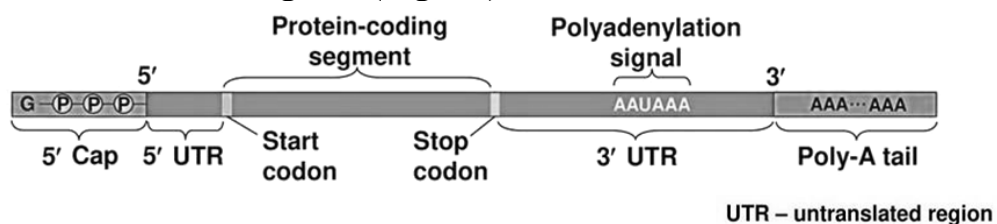


Figure 51. The structure of eukaryotic mRNA

tRNA

- Transfer RNAs (tRNAs) are non-coding RNAs that transport amino acids into the ribosome and ensure that they correspond to the nucleotide sequence of mRNA. In essence, they are translators of nucleic acid sequence language into amino acid sequence language.
- A ribosome is a complex of proteins and rRNA, which ensures the correct spatial arrangement of the other participants of translation and catalyzes the incorporation of a new amino acid into the polynucleotide sequence.
- Regulatory proteins that contribute to both the initiation and the course of translation (initiation factors, elongation factors).

In a two-dimensional diagram, tRNA resembles a cloverleaf with several stems. These stems are formed through complementary interactions between

different sites within the same nucleotide chain. One of the tRNA stems consists of the 5' and 3' ends joined together and is called the acceptor stem. The amino acid corresponding to this tRNA binds to the adenosine at the 3' end.

At the opposite end of the molecule is an **anticodon**, a triplet of nucleotides that determines which amino acid can be bound to the given RNA. For example, tRNA with a 5'-GAA-3' anticodon can only carry the amino acid phenylalanine. An anticodon is a three-nucleotide sequence that can bind complementarily to a particular codon in mRNA. Such binding occurs in an antiparallel manner. For example, the 3'-AAG-5' anticodon binds to the mRNA 5'-UUC-3' codon encoding phenylalanine. This is how the complementarity of mRNA and tRNA sites ensures that specific codons correspond to specific amino acids (Fig. 52).

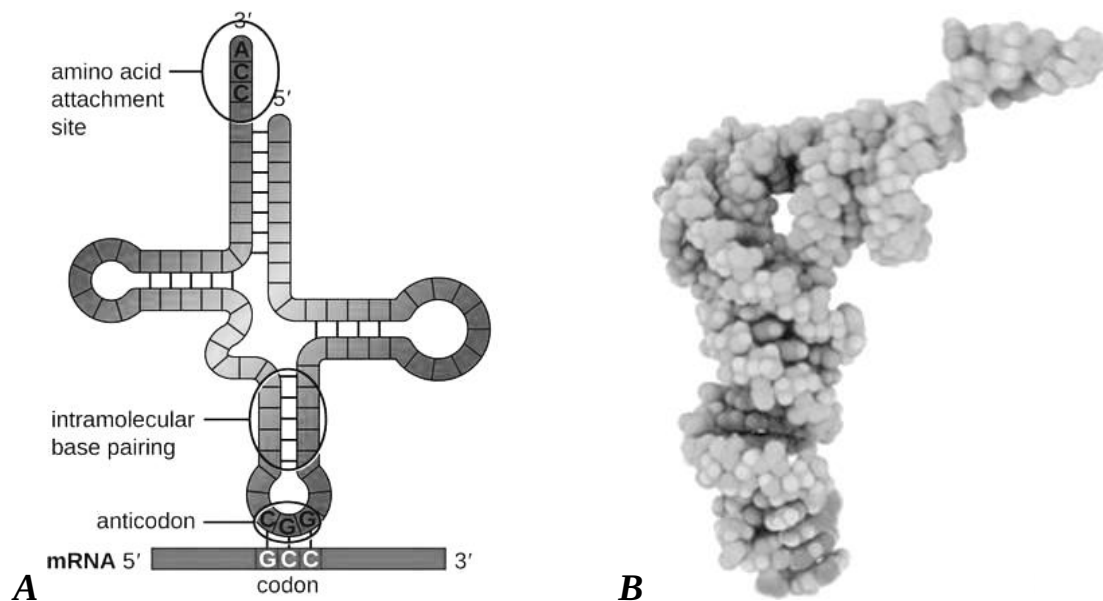


Figure 52. The structure of tRNA:
A — two-dimensional diagram; B — three-dimensional model

Recognition of tRNAs and attachment of their corresponding amino acids are provided by the enzymes aminoacyl-tRNA synthetases. The number of these enzymes corresponds to the number of proteinogenic amino acids. Each of them recognizes both its amino acid and all of its corresponding tRNAs. Aminoacyl-tRNA synthetase attaches the amino acid to the tRNA forming an **aminoacyl-tRNA**, a transfer RNA carrying the amino acid. This binding is ATP-consuming, so the bond between the nucleotide and the amino acid is high-energy, and the aminoacyl-tRNA becomes activated.

RIBOSOME

Recall that the ribosome consists of rRNA and proteins (see Chapter 2). Peptidyltransferase, a ribosome enzyme that forms peptide bonds, is also an RNA.

The ribosome consists of large and small subunits; the large subunit contains three sites through which each aminoacyl-tRNA sequentially passes during translation (Fig. 53):

- The **A-site** is the aminoacyl site. It is where new aminoacyl-tRNAs arrive.
- **P-site** — peptidyl site. During translation, it constantly contains peptidyl-tRNA, tRNA, carrying not just one amino acid, but a chain of them (peptide).
- The **E-site** is an exit site. The tRNA freed from amino acids lingers there for some time before leaving the ribosome.

The position of the ribosome on the mRNA is such that the E site faces the 5' end and the A site faces the 3' end (the ribosome moves towards the 3' end of the mRNA). Thus, the direction of tRNAs passing through the A, P, and E sites, in turn, is opposite to the direction of ribosome movement.

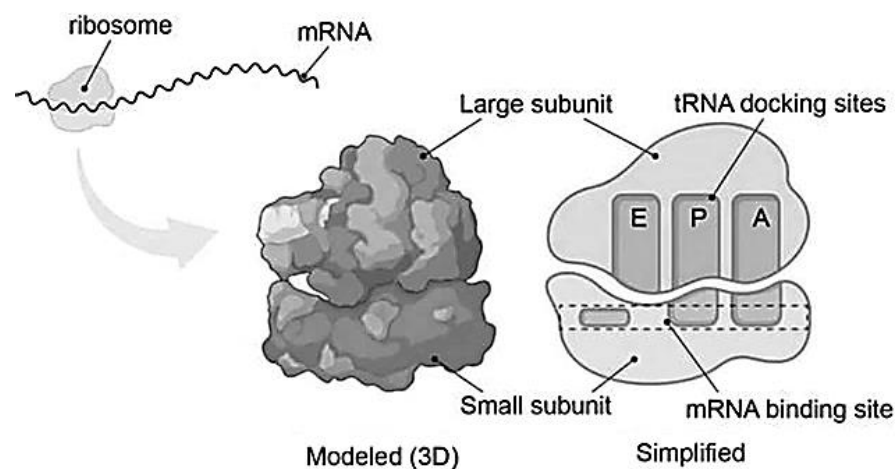


Figure 53. Diagram of the ribosome structure

TRANSLATION

Initiation. Translation initiation is a complex process in which the first tRNA and ribosomal subunits assemble into a single complex at the mRNA site containing the start codon. This occurs with the help of special proteins. First, the complex of mRNA, the first tRNA, and the small ribosomal subunit is formed, and then the large subunit is attached. As a result, the first tRNA immediately ends up in the P site (the A site remains empty). In prokaryotes, the small subunit recognizes a special mRNA nucleotide sequence, which is complementary to the 3'-end of the small rRNA subunit.

In eukaryotes, there is no such sequence, and the small subunit initially binds to the 5' cap and then moves along the mRNA until it reaches the first start codon in its path (Fig. 54).

Elongation. Elongation is a stage of translation in which new amino acids are added to those already present, resulting in the elongation of the polypeptide molecule. This process also occurs with the participation of auxiliary proteins,

and elongation factors. This process consists of repeated steps: The aminoacyl-tRNA enters the free A-site.

Eukaryotic translation initiation

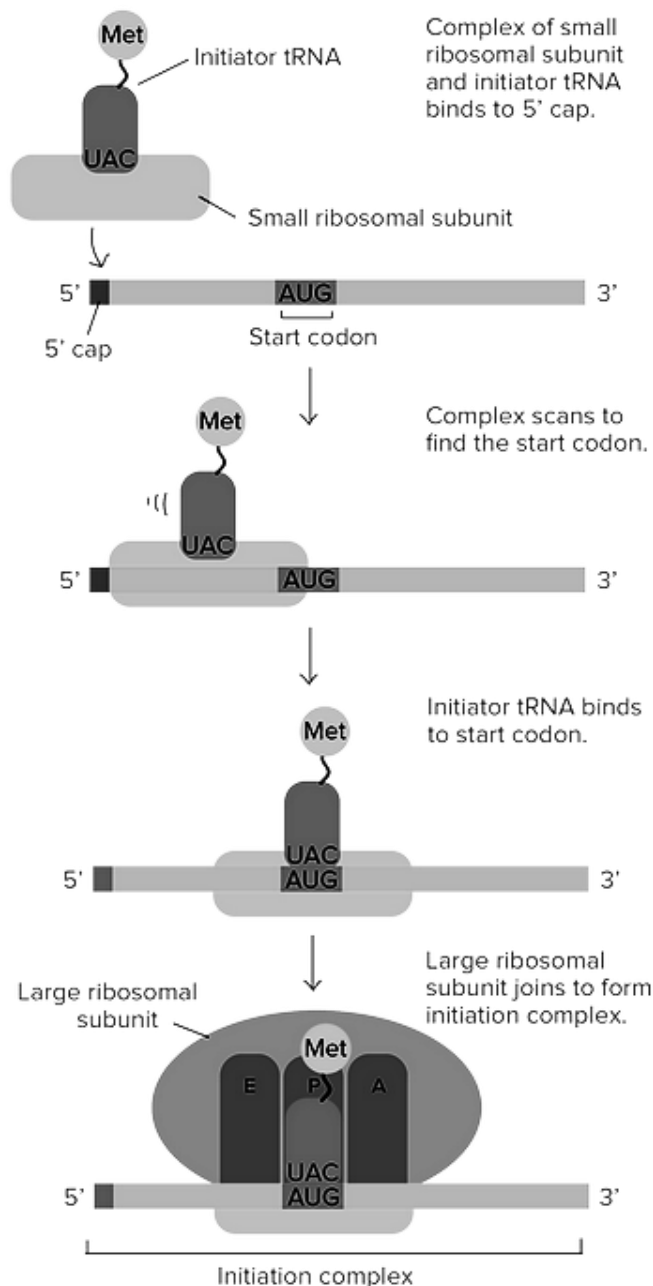


Figure 54. Initiation of translation in eukaryotes

begins to protrude from the ribosome through a special channel. If the amino acid chain emerging from the ribosome contains a special sequence, it is recognized by a molecular complex called the **signal recognition particle** (SRP). As a result, such a ribosome attaches itself to the ER, translation continues, and the synthesized peptide enters the lumen of the ER through a special protein channel (**translocon**). If the synthesized peptide does not have such a sequence, the ribosome remains unbound to the ER.

If the mRNA codon in the A-site matches the tRNA anticodon (i.e., binds complementarily to it), a peptide bond is formed between the amino acid delivered to the A-site and the amino acid contained on the tRNA in the P-site.

A peptide bond is formed, with the amino acid that was attached to the tRNA in the P-site being detached from it and attached to the "new" amino acid bound to the tRNA delivered to the A-site.

After the peptide bond is formed, the ribosome moves one codon to the 3' end of the mRNA. As a result: The next codon along the mRNA to the 3' end enters the A-site.

The transfer RNA to which the amino acids of the resulting polypeptide (peptidyl-tRNA) are linked moves from the A-site to the P-site. The "empty" tRNA that was previously in the P-site is now in the E-site and will subsequently be removed from the ribosome (Fig. 55).

The repetition of these steps leads to the growth of the polypeptide. As a result, it

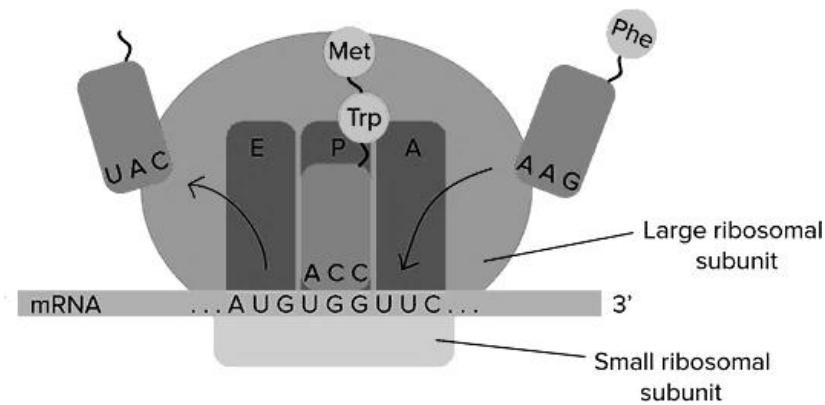


Figure 55. Elongation

Termination. Elongation continues until a stop codon enters the A-site. It does not encode amino acids but is recognized by special proteins — termination factors — proteins that release the newly created polypeptide from the ribosome (Fig. 56).

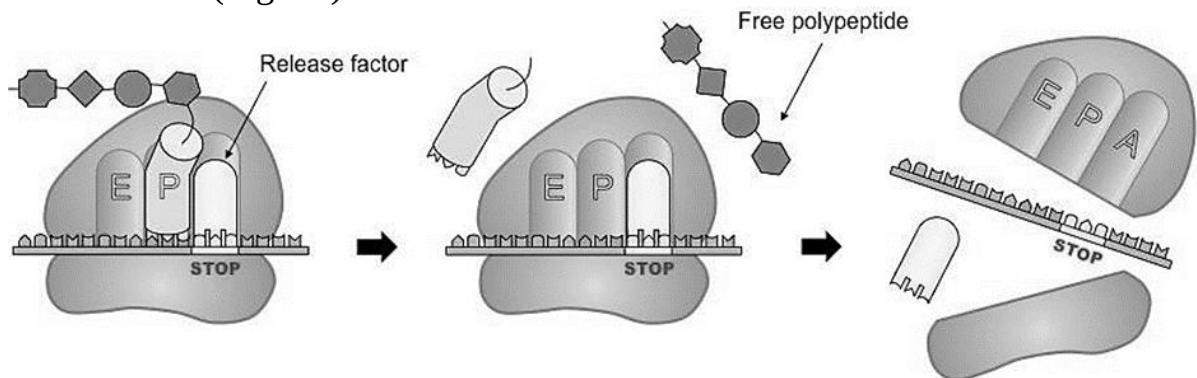


Figure 56. Termination of translation

PROTEIN FOLDING AND POST-TRANSLATIONAL MODIFICATIONS

There are four main levels of protein molecule organization: **primary**, **secondary**, **tertiary**, and **quaternary** structure. The primary structure is a chain of amino acids connected by peptide bonds. The other levels of the organization are provided by interactions between amino acids of the polypeptide chain, which creates a unique three-dimensional structure from a linear molecule. This structure is called the native conformation of the protein. The process of a protein acquiring its native conformation is called **folding**. Only polypeptides that have acquired the necessary structure become functional proteins. Therefore, the development of sufficiently reliable software to predict the structure of a protein from its amino acid sequence is of great importance for proteomics.

Protein folding occurs spontaneously and has various driving forces. An important mechanism is an interaction with the aqueous environment in which the protein exists. In addition, protein folding occurs due to the formation of hydrogen and other bonds, Van der Waals forces. This process often occurs during the translation with an already synthesized protein fragment.

This process also depends on environmental conditions such as salt concentrations, pH, and temperature. If these conditions deviate significantly from optimal ones, even the native protein can lose its structure. This phenomenon is called **denaturation**.

The concentration of proteins in the cytosol is very high. Therefore, there is also a high probability that the emerging protein will collide with other proteins before its folding is complete. As a result, their exposed hydrophobic sites can combine and lead to the formation of an insoluble protein aggregate. Thus, a large proportion of the proteins produced in the cell could be dysfunctional and subject to disposal. In reality, however, more than 95 percent of the proteins present in cells are in their native conformations. This is due to the presence of special proteins called chaperones in the cells of all organisms. Chaperones are proteins that ensure the proper folding of other proteins (Fig. 57). Chaperones are present in every cellular compartment and organelle. In some cases, improper protein conformation can be the cause of the development of severe diseases. There are proteins that, having the wrong conformation, can give it to other similar proteins. Such proteins are called prions. Prions form abnormal clusters of proteins in cells, causing cell death and tissue damage. Prion diseases are rare diseases of the nervous system that occur not only in humans but also in other animals. An example is Creutzfeldt-Jakob disease (CJD), which affects the cerebral cortex and other parts of the brain, leading to death in 100 % of cases. In cattle, the disease is called mad cow disease.

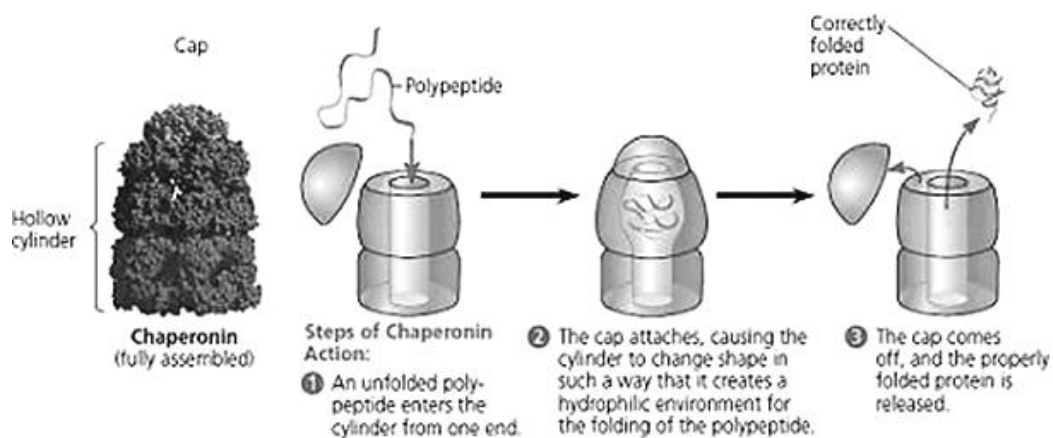


Figure 57. Protein folding scheme using chaperonin

Post-translational modification (PTM) is a biochemical modification with one or more amino acids in a protein after it has been synthesized by the ribosome. Almost all proteins undergo it. This process can occur at any stage of a protein's life. The same protein can undergo several different modifications. Many modifications occur shortly after translation is completed and may ensure proper protein folding or serve to direct the protein to certain cellular structures (e.g., nucleus or membrane). Other modifications occur after folding and localization are complete to activate or inactivate catalytic activity.

Several hundred types of post-translational modifications are currently known. Among them: hydrolysis of peptide bonds, formation of disulfide bonds, attachment of chemical groups (phosphorylation, glycosylation, acetylation, lipid attachment, ubiquitination). Various types of posttranslational modification of proteins can occur in the ER, Golgi apparatus, and elsewhere in the cell.

Many pathological conditions can be associated with a lack of post-translational modification of proteins. For example, the main protein of connective tissue collagen needs modification by hydroxylase enzymes. These enzymes can function only with vitamin C (ascorbic acid) in their composition. Therefore, in the case of scurvy (a disease resulting from a lack of vitamin C), insufficiently hydroxylated collagen does not form normal fibrils. Without the structural support of this protein, blood vessels, tendons, and skin become brittle, which accounts for the symptoms of this pathology. Mutations in the gene for one of the hydroxylases can also cause connective tissue defects.

DEGRADATION OF PROTEINS IN THE CELL

The lifespan of proteins varies from a few minutes (cyclins) to the lifespan of an organism (lens proteins). It is controlled by mechanisms of **protein degradation** — i.e., their breakdown into amino acids. These mechanisms perform several biological functions. The first is the removal of potentially dangerous proteins (those with the wrong conformation, damaged proteins). The second is the maintenance of a certain level of a particular protein concentration in the cell or its modification in response to changes in the conditions of existence.

Eukaryotic cells have several pathways for protein degradation. One of the main pathways is hydrolysis in **lysosomes**. It is mostly directed to old or defective cell organelles and extracellular structures absorbed during phagocytosis. Another pathway is protein degradation in **proteasomes**, special protein complexes designed to break proteins. Up to 90 % of proteins in mammalian cells are utilized in this way. This process occurs with energy expenditure.

This mechanism of protein molecule utilization is called **ubiquitin-dependent protein degradation** (Fig. 58). To be sent to the proteasome, a protein must have a special “tag” — another small protein called **ubiquitin**. The attachment of ubiquitin to other proteins is called ubiquitination and is performed by ubiquitin ligase enzymes.

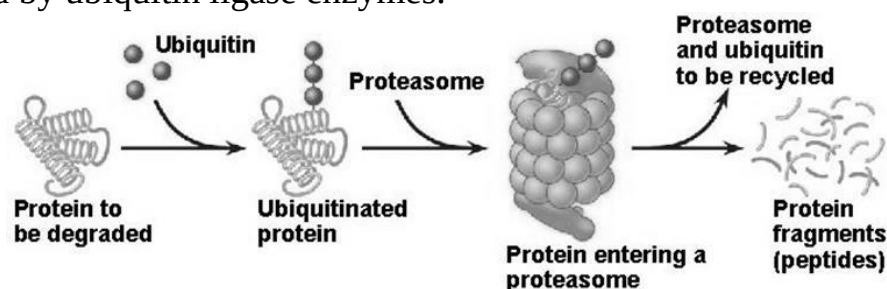


Figure 58. Ubiquitin-dependent protein degradation

CHAPTER 6

REGULATION OF GENE EXPRESSION

CLASSIFICATION OF GENES

All cells of a multicellular organism with a nucleus contain the same set of genes. However, cells of different tissues (and even cells of the same tissue) can be very different from each other. It is provided by the *expression of different genes*. The expression of different sets of genes in the cells of the same organism is called **differential gene expression**. There are housekeeping genes and tissue-specific genes.

Housekeeping genes encode proteins necessary to maintain basic cell functions and must be expressed in all cell types. They include genes encoding proteins that are involved in transcription and translation, ribosome and proteasome proteins, histones, many cell cycle regulators, enzymes involved in the most important catabolic and anabolic reactions, lysosomal proteins, cell membrane proteins (including transport proteins), various signaling proteins and many others.

Tissue-specific genes are genes whose products are responsible for individual characteristics specific to a particular cell type. They are significantly expressed only in cells of one or more tissue types.

A typical human cell can express from 1/3 to 1/2 of its protein-coding genes at a time. Highly differentiated cells, such as muscle or nerve cells, express a smaller portion of their genes. Also, another classification divides genes into functional (encoding proteins or RNA aimed at regulating other genes) and structural (encoding proteins and RNA that are not regulatory).

REGULATION OF GENE EXPRESSION

Gene expression refers to the process of converting information stored in DNA into protein or, in some cases, into non-coding RNA. Simplifying the concept greatly, we can say that if a gene is an “instruction for assembling a protein” then gene expression is “assembling a protein according to the instruction.”

All organisms, both unicellular and multicellular, regulate the expression of their genes. The process of regulation can take place at several levels. In eukaryotes it can occur:

- at the level of chromatin rearrangements (changing the location of nucleosomes to allow access of regulatory proteins to genes);
- at the level of transcription initiation — transcription starts only under certain conditions;
- post-translation — some mechanisms affect certain mRNAs and affect protein synthesis.

TRANSCRIPTION REGULATION IN PROKARYOTES

Due to the simpler organization of the prokaryotic genome, they mainly regulate gene expression at the *level of transcription initiation*. This is done by special regulatory proteins that bind to specific DNA sequences. As a result, *the binding of RNA polymerase to the **promoter*** (the site where RNA polymerase binds to DNA and starts transcription) is either *stimulated (positive regulation)* or *inhibited (negative regulation)*. Proteins that stimulate transcription are called **activators** and those that inhibit it are called **repressors**.

OPERON

If several proteins of a prokaryotic cell are members of the same metabolic pathway, they are often located together in the genome and form a unit of transcription that is called an **operon** (Fig. 59). Operon is a group of genes whose transcription occurs from the same promoter and is controlled by the same operator. An **operator** is a small region of DNA located near the promoter. A special protein can bind to the operator and prevent transcription. This protein is called a **repressor**. Each operon in a bacterium is regulated by a specific type of repressor. All genes of the same operon are transcribed into the same mRNA. As a result, such mRNA contains several protein-coding sequences, each with its start and stop codons. These sequences are called **cistrons**, and mRNA containing more than one cistron is called **polycistronic**.

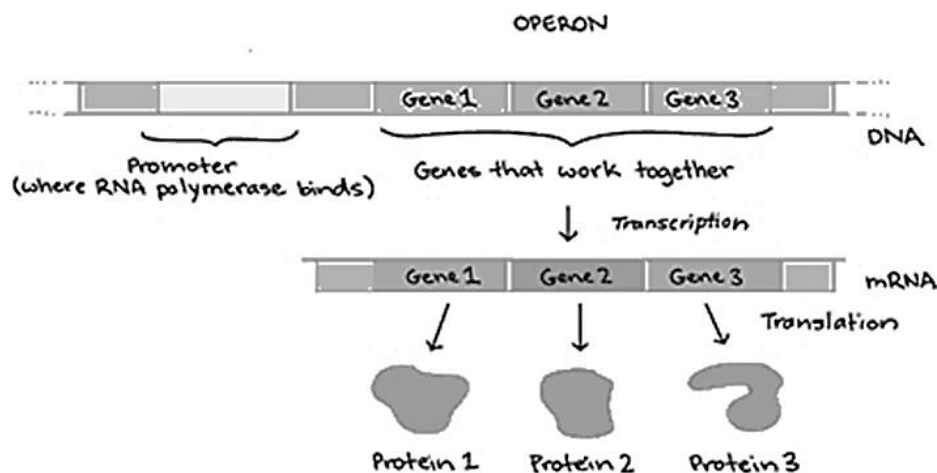


Figure 59. Diagram of the operon

Changes in the level of gene transcription in response to the chemical composition of the environment in bacteria can be roughly divided into two types: a substance either stimulates the production of a certain protein or, on the contrary, inhibits it. This is called **regulation by induction** and **regulation by repression**. The substance that induces transcription is called an **inducer**, and the substance that prevents it is called a **co-repressor**. Both mechanisms can be illustrated by two operons present in many bacteria: the *lac operon* and the *trp operon*.

Lac operon. The **lactose operon (lac operon)** is a group of genes that allow a bacterial cell to feed on lactose. One of them is the gene for the enzyme β -galactosidase, which cleaves this disaccharide to glucose and galactose. Active transcription of these genes can start only in the presence of lactose. If lactose is added to the medium, the number of β -galactosidase molecules increases by 1000 times in about 15 minutes.

The following elements are distinguished in the Lac operon (Fig. 60):

- **Promoter** and **terminator** — the sites where RNA polymerase binds to DNA and where it dissociates from it.
- **Genes** involved in lactose metabolism. They are located between the promoter and terminator. One of them encodes the enzyme β -galactosidase.
- The **operator** is located near the promoter. It is a nucleotide sequence that is recognized by the *repressor protein*, which prevents transcription.

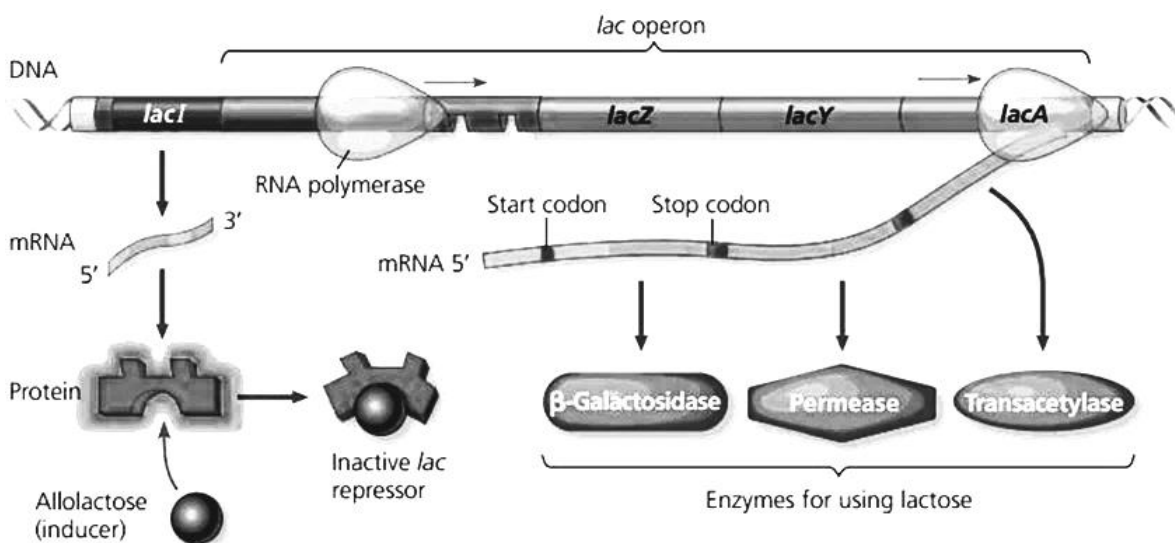


Figure 60. Lac operon

The repressor plays a key role in lac operon regulation. Let us consider this process in more detail:

- Since the repressor is a protein, it is also encoded by a certain gene, but this gene is not part of the operon itself (**regulatory gene**).
- Small amounts of repressor are constantly synthesized in the bacterial cell.
- If the repressor is bound to the operator, the RNA polymerase has no access to the promoter and no transcription takes place.

Repression of transcription in the lac operon by the repressor protein:

- The repressor is designed so that lactose binds to it and changes its conformation. As a result, the *repressor becomes unable to bind to the operator*, and RNA polymerase can transcribe structural genes.
- If all the lactose is utilized, new repressor molecules will not bind to it and therefore will remain capable of binding to the operator and blocking transcription.

Thus, the *lac* operon is regulated by induction since lactose induces transcription. In this case, *lactose is called an inducer* (Fig. 61).

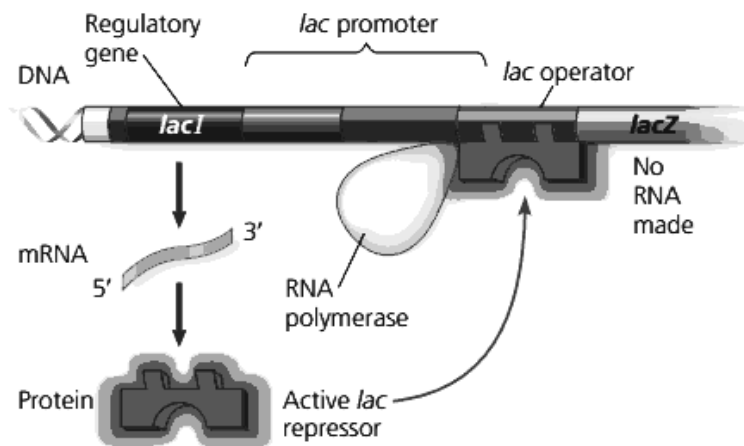
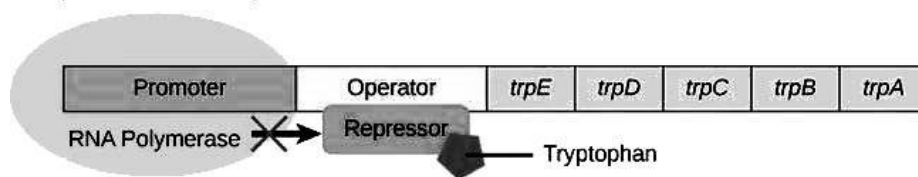


Figure 61. Negative regulation of *lac* operon by a repressor protein

Trp operon. The **tryptophan operon (*trp* operon)** is a group of genes that encode proteins necessary for the synthesis of the amino acid *tryptophan* (Fig. 62). When this amino acid is present in the environment, there is no need to produce these proteins. Therefore, tryptophan inhibits the production of proteins and is a co-repressor.

When tryptophan is present, the *trp* repressor binds the operator, and RNA synthesis is blocked.



In the absence of tryptophan, the repressor dissociates from the operator, and RNA synthesis proceeds.

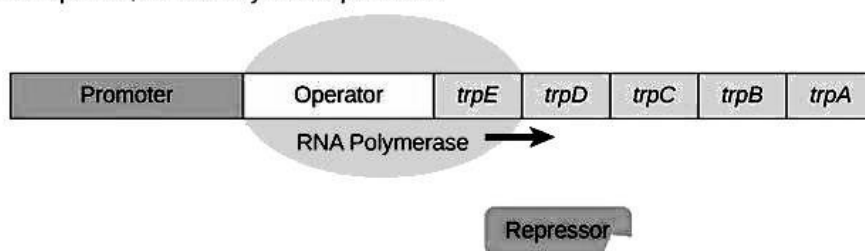


Figure 62. Diagram of *trp* operon

The *Trp* operon can also be viewed as a segment of DNA on which are located:

- Promoter and terminator.
- Between the promoter and terminator are five structural genes. The proteins they encode are involved in tryptophan synthesis.
- Operator — located near the promoter.

The mechanism of negative regulation of the trp operon is similar to that of the lac operon.

- A repressor can bind to the operator. The repressor of this operon is also encoded by a gene that is not part of it.
- Unlike the lac operon, which is regulated by induction, the trp operon is regulated by repression: the repressor can join the operator only if it binds to tryptophan (co-repressor).
- If there is tryptophan in the medium, no transcription occurs, but in its absence, the repressor is no longer active, and the production of proteins required for tryptophan synthesis begins.

REGULATION OF TRANSCRIPTION IN EUKARYOTES

Transcription in eukaryotes is regulated by more complex mechanisms. This is because in eukaryotes DNA is organized into **chromatin** and, due to the presence of the nucleus, transcription, and translation are separated in space and time, which opens up opportunities for a greater diversity of regulatory mechanisms. In recent decades, the high importance of *epigenetic mechanisms* of gene expression regulation has been realized, which will be discussed below.

Transcription initiation in eukaryotes. As in bacteria, the initiation of transcription in eukaryotes is controlled by proteins that bind to DNA and either promote or prevent RNA polymerase binding. In eukaryotes, however, this process is more complex.

In eukaryotes, RNA polymerase recognizes the promoter with the help of special proteins called **transcription factors (TF)**. There are general and specific TF. **General transcription factors** are the proteins that are involved in the transcription of all protein-coding genes. They are essential for the initiation of transcription. The protein complex on the promoter that consists of general transcription factors and RNA polymerase II is called the **preinitiation complex**.

Specific transcription factors, unlike general ones, are not obligatory. Specific transcription factors are involved in the transcription of only certain genes. Their role is to change the level of expression of these genes: they either stimulate it (activators) or inhibit it (repressors). The action of various specific transcription factors is determined by the type of cell or by certain signals, which can increase the rate of gene expression.

The promoters for RNA polymerase II can have several different types of sequences that are key to transcription initiation. Such sequences include TATA boxes, initiators, and CpG islands.

Enhancers and silencers. Some DNA regions recognized by transcription-regulating proteins are located close to the promoter. However, other regulatory elements can be located in thousands of base pairs upstream or downstream from the promoter, even within introns. They are called **enhancers** and **silencers**. They can be bound with the proteins (activators and repressors),

which can interact with the preinitiation complex (Fig. 63). Interactions of proteins located at such distant sites of DNA are possible because DNA forms loops that can be 1 million base pairs long. By binding activators or repressors, the rate of gene expression in eukaryotes can be greatly increased or decreased. A gene may have several enhancers or silencers, each of which may be active at different times and in different cell types. A combination of these regulatory elements is important to achieve a certain effect on transcription.

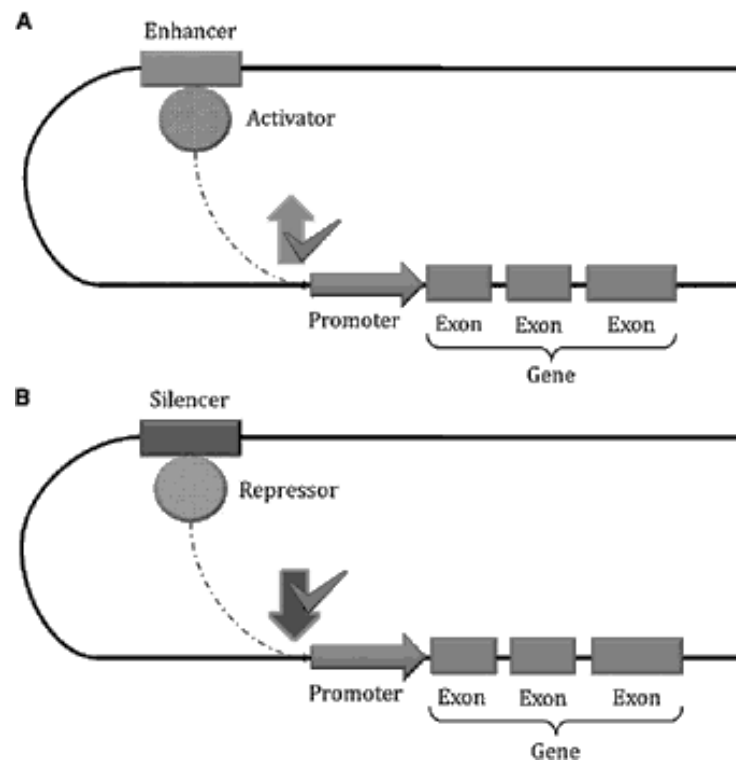


Figure 63. Diagram of enhancer (A) and silencer (B)

STRUCTURAL ORGANIZATION OF THE NUCLEUS AND GENE EXPRESSION

Chromosomes in the nucleus are arranged not randomly but occupy their **chromosome territories**. Studies show that transcriptionally active genes are located at the borders of these territories. This organization is believed to coordinate the expression of these genes. There are also regions of active transcription which are called **transcription factories**. There is evidence that some factories transcribe the genes regulated by the same activators. Thus, it is known that the structural organization of chromatin in the nucleus influences the regulation of transcription. It is a subject of active research.

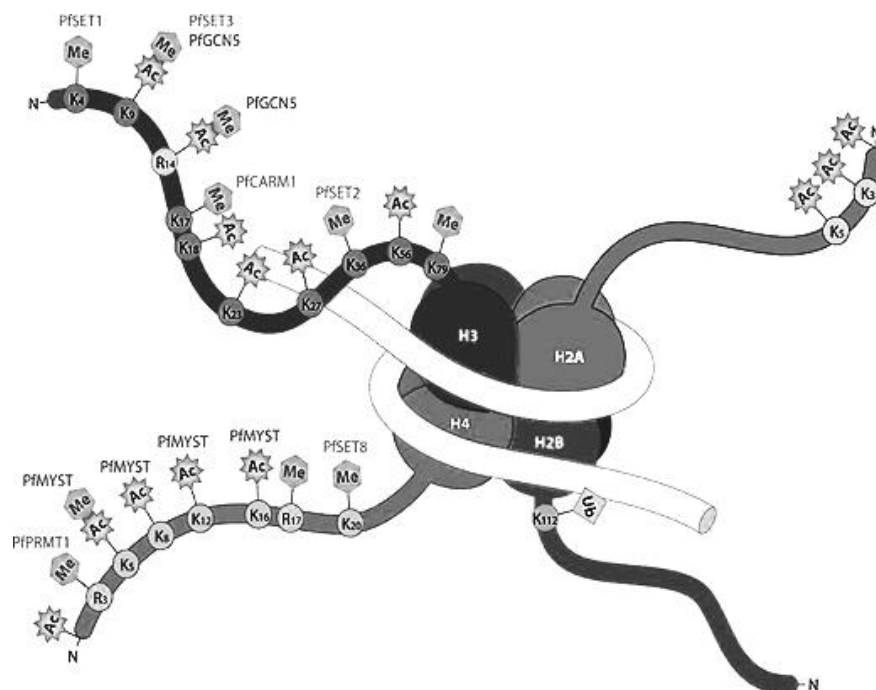
HISTONE MODIFICATIONS AND CHROMATIN REMODELING

Eukaryotic DNA is bound with octamers of histone proteins, forming **nucleosomes**. This allows DNA to be tightly “packed” in the nucleus. A diploid human cell has about 3×10^7 nucleosomes. Another function of

histones is participation in the regulation of gene expression. Genes located in heterochromatin regions are usually not expressed. This is because the proteins required for transcription do not have access to the regions of DNA tightly bound to histones and other chromatin-binding proteins. The Chromatin of this conformation is often referred to as “closed.” For transcription to become possible, the chromatin must become “open” to RNA polymerase. The process by which histone-bound DNA becomes accessible to regulators that are necessary for transcription is called **chromatin remodeling**. It is also one of the mechanisms of the epigenetic inheritance of traits.

There are several ways to obtain “open” DNA:

- **Modification of histones.** Histones have so-called **tails** — terminal regions of a polypeptide that protrude outside the nucleosome. Special regulatory proteins can covalently attach some functional groups to them. The most common modifications of histones are the addition of **acetyl** ($-\text{COCH}_3$), **methyl** ($-\text{CH}_3$), or **phosphate** groups (Fig. 64).



Several amino acids of the histone tail can be modified simultaneously. The combination of such “tags” can have different effects, such as attracting certain regulatory proteins. It can both promote transcription and suppress it. However, some common patterns can be noted: for example, histone acetylation

more often promotes transcription, while methylation is the opposite. To illustrate this mechanism, consider the following example: methylation of the ninth lysine in the tail of histone H3 attracts a protein that binds this nucleosome to another, and, as a result, a more compact “packing” of chromatin occurs.

- The **rearrangement or removal of nucleosomes from DNA**. This process is carried out by large multisubunit enzymes that use the energy of ATP hydrolysis. For example, they can loosen the bond between histones and DNA, causing the nucleosome to “slide” over DNA, opening up regulatory regions on it.

DNA METHYLATION AND METHYLOME

Not only histones but also DNA itself can be modified. Common modifications are **methylation and demethylation of DNA** — attachment or removal of methyl groups to nitrogenous bases. In humans, methylated cytosine is present in about 1.5 % of genome DNA. Usually, such methylated cytosine is found in “CG” dinucleotide sequences (in **CpG islands**, Fig. 65) in the promoter regions of many genes (in humans, about 70 % of promoters).

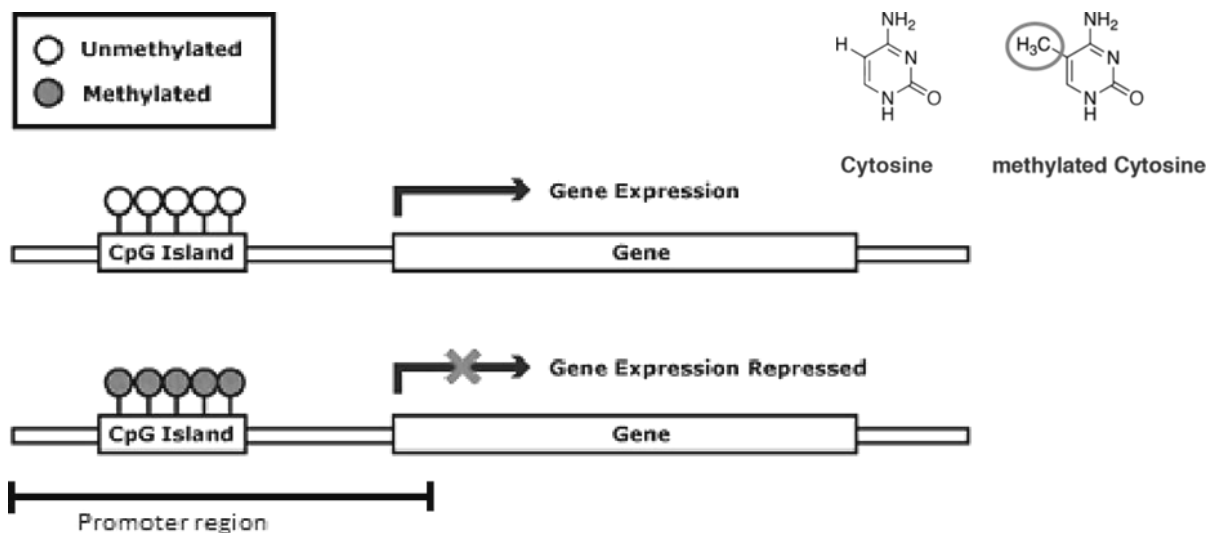


Figure 65. DNA methylation in promoters of genes

Methyl groups prevent the binding of transcription factors and other proteins necessary to initiate transcription. Thus, methylation is a way to “turn off” the gene.

A particular nucleic acid methylation “pattern” present at a particular time in the genome or a particular cell type is called **methylome**. Methylome depends on the cell and tissue, and is not constant; it changes as cells respond to conditions of existence. In mammals, DNA methylation occurs after DNA replication and during cell differentiation, and methylome is usually inherited by daughter cells and therefore is one of the mechanisms of epigenetic inheritance of traits.

RNA INTERFERENCE

RNA interference (RNAi) is the inhibition of the expression of a particular gene by a special RNA that binds to the complementary single-stranded mRNA encoded by that gene and either suppresses translation or causes mRNA degradation.

Noncoding RNAs involved in RNA interference are broadly referred to as small non-coding RNAs. They are double-stranded molecules 20–31 nucleotides long. They include *small interfering RNAs* (siRNAs) and *microRNAs* (miRNAs). They come from different sources, but their mechanisms of action are similar.

Small interfering RNAs are produced from longer double-stranded RNA molecules that can appear in cells as a result of viral infection or transposon expression. The resulting RNA then binds to **RISC (RNA-induced silencing complex)**. Upon binding to the mRNA, RISC cleaves it in the complementary match region with a small interfering RNA (Fig. 66).

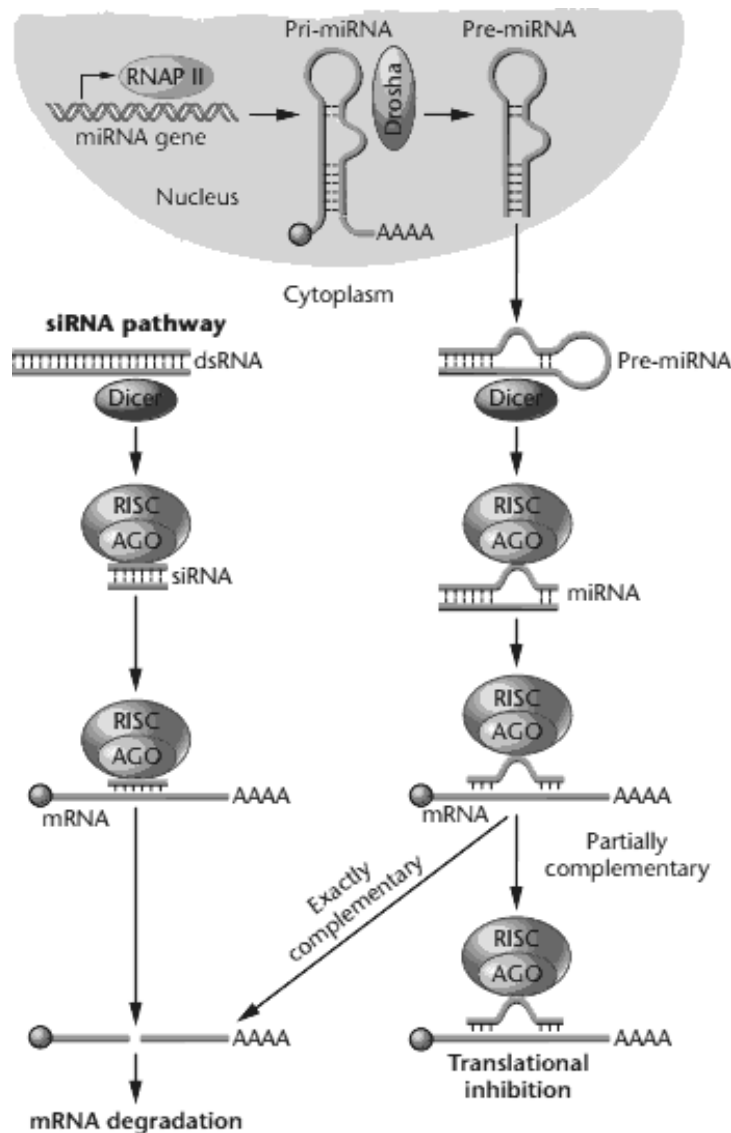


Figure 66. Pathways of microRNA and siRNA

MicroRNAs are found in plants and animals. They originate in the nucleus by transcribing from microRNA genes. Studies show that there are at least 1,500 such genes in the human genome. RISC-bound microRNAs target complementary sequences in the mRNA. If the complementary match of microRNA and mRNA is perfect, mRNA is cleaved by RISC and degraded. If the match is incomplete (which is typical for animals), partial suppression of translation occurs. Synthesis of microRNAs to suppress the activity of mRNAs created by the cell itself can provide rapid changes in gene expression, if necessary, since the suppression of transcription is usually insufficient to stop the production of a protein that is no longer needed. This is because mRNAs can exist in the cytoplasm for a relatively long time.

LONG NON-CODING RNAs

Eukaryotic genomes also encode many **long noncoding RNAs, lncRNA** (more than 200 nucleotides long). According to some estimates, the human genome encodes about 17,000 such RNAs. They are formed similarly to mRNAs (capping and polyadenylation occur, and splicing is possible), but do not have start and stop codons, indicating that they do not encode proteins. The regulatory functions of long noncoding RNAs are diverse and are actively investigated. Among them are the following:

- Some such RNAs influence the expression of various genes by binding to **chromatin remodeling complexes**.
- Some can regulate transcription by binding directly to transcription factors.
- Some can bind to complementary mRNAs or pre-mRNAs, which can affect alternative splicing and also lead to the formation of double-stranded RNA, which in turn triggers the mechanism of RNA interference.

Regulation at the translation level. Gene expression is also regulated at the translation level, most commonly at the initiation stage. This can occur with the help of regulatory proteins that bind to specific sequences or structures in untranslated regions at the 5' or 3' end, preventing ribosome attachment.

EPIGENETIC INHERITANCE

As we studied the organization of the genome and the regulation of gene expression, it became clear that the inheritance of some traits is not explained by the DNA sequence of the organism alone. An example is demonstrated by the gene *agouti* in mice, which determines the color of the coat. Normally, it is expressed in the developing organism only for a short time — when the coat is formed. But there were mice in which *agouti* was expressed throughout their development. Because of this, they had a yellow color that is not characteristic

of normal mice (as well as some other traits). However, if the diet of pregnant mice is supplemented with compounds containing **methyl groups** (e.g., folic acid), the color of the offspring's hair returns to normal, and this effect persists even in subsequent generations. Thus, it was found that the degree of color change correlates with the level of DNA methylation. The explanation for this phenomenon is the presence of a CpG residue upstream of the agouti gene. In the presence of sufficient sources of methyl groups, the gene is “turned off” during the required period of embryonic development and remains in this state throughout the animal’s life. Thus, the differences between the mice in this example are due to the nutrition of their mothers during pregnancy (Fig. 67).

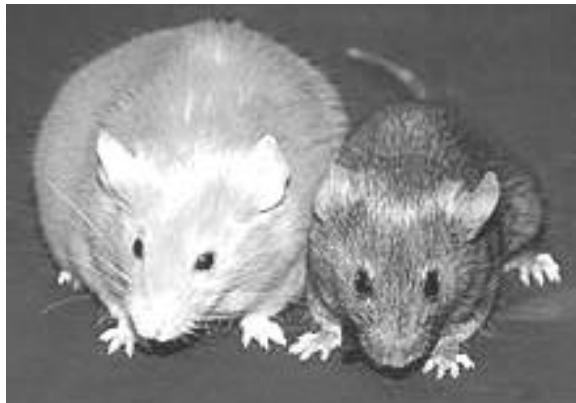


Figure 67. An agouti mouse and a normal mouse

A similar effect associated with changes in DNA methylation has been observed in humans. At the end of World War II, in the winter of 1944–45, there was a famine in the Netherlands. Over time, it was discovered that children of women who were in early pregnancy at the time had higher risks of developing obesity, elevated blood triglyceride, and cholesterol levels, type 2 diabetes, and schizophrenia as adults. A study comparing such people with their siblings born before or after the events described found a difference in the DNA methylation of certain genes. The peculiarities of the children born under the described conditions can be interpreted as follows: when the pregnant woman was undernourished, an epigenetic “tuning” of the metabolism took place in the developing organism, which prepared it for the subsequent existence in starvation conditions and persisted throughout its life. However, later on, after birth, it did not correspond to the real environmental conditions, since the amount of available food turned out to be sufficient. As a result, the metabolic peculiarities of such people turned out to be harmful rather than adaptive.

Similar mechanisms also explain the cases when one of the monozygotic twins, despite practically identical genomes, develops a hereditary disease, while the other does not. Changes in the methylome are observed in some forms of cancer, causing “incorrect” gene expression.

Studies of such phenomena led to the emergence of epigenetics, the study of changes in organisms caused by modification of gene expression rather than alteration of the genetic code itself. There are three main epigenetic mechanisms:

- Reversible modification of DNA by methylation and demethylation.
- Remodeling of chromatin.
- Regulation of gene expression by short and long non-coding RNAs.

The specific structure of the described epigenetic modifications present in the cell at a certain period is called an **epigenome**.

As mentioned earlier, the diversity of cell types in the organism is possible due to the expression of different sets of genes in them. This is also partly due to the differences in their epigenome. The epigenome is usually inherited by daughter cells during replication and mitosis. Thus, after each DNA replication cycle, methylation of the daughter strand occurs on sections of DNA where one strand is already methylated. This contributes to the fact that the daughter cells remain properly differentiated (i.e., specialized) cells of the same type as the mother cell.

However, the zygote from which the organism originates is not a specialized cell of any tissue. Because of this, during the formation of gametes, the existing DNA methylation “pattern” is largely erased and is then mapped differently in male and female germ cells. Thus, the fertilized egg cell has different “marks” on the copies of certain genes received from the mother or father. Soon after fertilization, during embryonic development, most of them are “erased” again, which will allow further creation of new epigenetic modifications “from scratch” to form more than 200 cell types found in the adult organism. However, some regions of the genome are not subjected to this re-demethylation, and the genes contained in them retain the “marks” of the maternal and paternal chromosomes. This can explain such a phenomenon as genomic imprinting — the different expression of an allele depending on the sex of the parent it was inherited from.

An example of imprinting is the inheritance of Prader–Willi and Angelman syndromes. If chromosome 15 with some defect in the long arm is inherited from the father, the child develops Prader–Willi syndrome, and if from the mother, Angelman syndrome. There are different hypotheses explaining the differences in gene methylation in female and male germ cells. The first one explains it by the difference in biological interests of the sexes: the sperm methylation pattern favors the embryo to receive as many resources as possible from the mother, while the oocyte methylation pattern favors the embryo to limit this influence. This is supported by the fact that most mammalian genes that undergo parental imprinting are in one way or another related to intrauterine development. Another hypothesis explains the imprinting by improving maternal-fetal compatibility: when some of the paternal genes are turned off, only maternal alleles will be expressed in the embryo, and

the embryo will become biologically and physiologically more similar to the mother's body. This may be supported by the fact that the number of "switched off" paternal genes is higher than the number of maternal genes, which finds confirmation.

CHARACTERISTICS OF THE HUMAN GENOME

The **genome** is the complete set of an organism's genetic information or the entire sequence of its DNA. It contains all the information an organism needs to function. The human nuclear genome is found in the chromosomes, which form 23 pairs. In addition to it, there is also the mitochondrial genome.

Human mitochondrial DNA consists of 16,569 base pairs and contains 37 genes. The nuclear genome consists of approximately 3,200,000,000 base pairs divided into 24 linear molecules (i.e., chromosomes). These 24 chromosomes include 22 autosomes and two sex chromosomes, X and Y.

Composition of the human genome (Fig. 68):

- Protein-coding genes About 20,000.
- Genes for long noncoding RNA About 16,000.
- Short non-coding RNA genes About 7,500.
- Pseudogenes About 15,000.

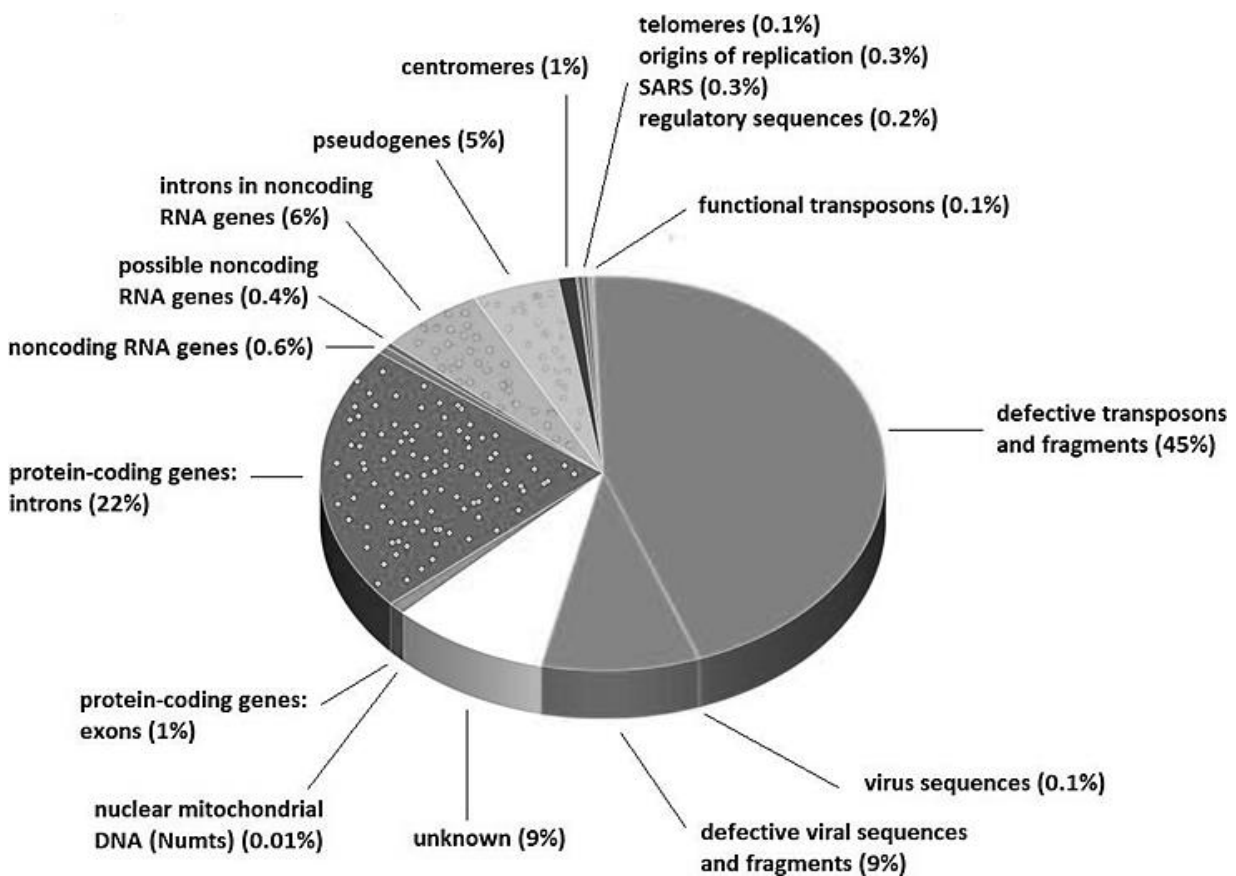


Figure 68. Human genome

The protein-coding part of the genome is small. From this fact, there appeared the idea that most of our genome is “junk DNA.” There is some indirect evidence to support this. However, as the ENCODE project has shown, about 80 % of the human genome can be transcribed, and large amounts of non-protein-coding RNA can be produced from the so-called “junk DNA.” Their functions have not yet been sufficiently studied.

GENOME REDUNDANCY AND ITS SIGNIFICANCE

The genomes of many organisms are also often characterized by some redundancy expressed in the fact that the biochemical function can be performed by the products of two or more genes. The main source of genetic redundancy is the process of gene duplication, which creates multiplicity in the number of their copies. A second and less frequent source of genetic redundancy is evolutionary processes leading to the appearance of genes that are similar in function but different in sequence.

- From an evolutionary point of view, gene duplications and subsequent division of functions between “copies” is a way of evolution of new genes, which, in turn, may contribute to the complexity of organisms. For example, the genes for all types of hemoglobin as well as for myoglobin originate from a single ancestral gene.

- From the aspect of ontogenesis, the presence of several “spare” genes performing similar functions can provide greater resistance of the developing organism to external factors that can negatively affect this process.

- In some cases, the increase in the number of gene copies can be explained by the need to produce a larger amount of the product of a given gene. This can explain the presence of about 300 copies of ribosomal RNA genes in several chromosomes in the human genome, which is about 0.4 % of the genome.

Genome size does not strictly correlate with organism complexity. The same applies to the number of genes; for example, both humans and the roundworm *C. elegans* have about 20,000 genes, but the complexity of the human organism is incomparably higher than that of this nematode. The lack of a strict correlation between the size of the genome and the complexity of its structure was called the **C-paradox**, and the discrepancy between the complexity and the number of genes was called the **G-paradox**.

REPETITIVE SEQUENCES

The human genome contains both **unique sequences** that appear only once in it (for example, sequences of protein-coding genes) and **repetitive sequences** — regions of DNA represented by several or a large number of

copies. According to their nature, repeats can be divided into **frequently repetitive sequences** (their number can reach one million copies per genome) and **middle repetitive sequences**. There are **tandem repeats**, which are sequenced one after another, and **interspersed repeats**, which are spread out over the genome and are spaced apart from each other.

Tandem repeats include:

- **Satellite DNA** — Frequently repeated tandem sequence most of which is the sequence of centromeres and telomeres.

- **Minisatellites** are a series of repeated fragments that are 10–100 base pairs long. The number of repeats of these sequences varies greatly from person to person, so they are *widely used in criminalistics, personal identification, paternity testing, and other areas*.

- **Microsatellites** are widely spread tandem repeats of short DNA sequences, usually consisting of 1–4 base pairs. There are **genes represented by multiple copies** (e.g., rRNA genes).

Interspersed repeats make up about 45 % of the entire human genome. Most of them are represented by varieties of **retrotransposons** called **LINE** and **SINE**.

DNA TRANSPOSONS AND RETROTRANSPOSONS

About half of the human genome consists of **mobile genetic elements** — DNA sequences capable of moving from one part of the genome to another (this process is called **transposition**). These include **DNA transposons** (also called **transposons**) and **retrotransposons**. A typical mobile genetic element resembles a simplified version of a virus.

In a simplified form, **DNA transposons** are regions of DNA flanked by special sequences. These regions contain the gene of the enzyme **transposase**. The enzyme recognizes the transposon, cuts it out, and inserts it into another part of the cell genome.

Retrotransposons also have repeats marking their boundaries and contain genes of **reverse transcriptase** and **integrase**. After transcription and synthesis of these enzymes, reverse transcriptase synthesizes a DNA copy of the retrotransposon RNA. Then, integrase inserts it into the cell genome. The most common types of retrotransposons in the genome are LINE (long interspersed nuclear elements) and SINE (short interspersed nuclear elements), which, according to some estimates, account for at least 34 % of the human genome. Only a very small portion of the LINEs are active. SINEs are relatively short non-coding retrotransposons. Because of this, they are not capable of independent transposition, and it is thought to be carried out at the expense of LINE retrotransposon enzymes.

HUMAN GENETIC POLYMORPHISM

All human DNA is structurally different from other human DNA. Differences in DNA sequences in individuals, groups, or populations are called **genetic polymorphisms**. The source of polymorphisms is mutations. Polymorphisms are present in both protein-coding and other parts of the genome.

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in humans. They represent a difference of one base pair in the DNA sequence of individual members of a species. Many of them (but not always) cause hereditary diseases such as sickle cell anemia or cystic fibrosis. The prevalence of SNPs in the genome is close to 1 : 1000 pairs of nucleotides. In addition to SNPs, individual human genomes have other differences associated with **deletions** (losses), **duplications** (doublings), or **inversions** (180° reversals) of sections of genomic DNA, as well as **copy number variations (CNVs)** of different sequences. The determination of individual differences, such as those of certain CNVs, is the basis of methods for the diagnosis of hereditary diseases, which will be discussed in more detail in the chapter “Human genetics.”

INTERNATIONAL HUMAN GENOME RESEARCH PROJECTS

Human genome project. The main goal of the Human Genome Project was to read the complete nucleotide sequence of the human genome, as well as to identify and map all of its genes. This international project began in 1990 and was completed in 2003. The project resulted in approximately 92 % of the genome sequencing of several donors. The development of DNA sequencing technologies has made this process somewhat faster and cheaper. The first reading of the human genome took more than 10 years and \$3.5 billion, but today advances in molecular biotechnology allow a single human genome to be read in just a few days and for a few tens of dollars.

Encyclopedia of DNA Elements (ENCODE). In 2003, the U.S. National Human Genome Research Institute launched the ENCODE project. The goal of the project was to identify all the functional DNA sequences that are found within the human genome, and subsequently the mouse genome, and to determine how these elements regulate gene expression.

Roadmap. The goal of the U.S. National Institutes of Health Roadmap Epigenomics Mapping Consortium is to create a catalog and public database of human epigenomic data that can be used for basic biological research as well as research focused on the pathogenesis of hereditary diseases. An epigenome is a set of specific tags that determine gene activity but do not affect the primary structure of DNA.

Human transcriptome, proteome, and metabolome. The human phenotype is determined not only by the presence of certain DNA sequences,

but also by the products of these sequences: RNA, proteins, and, ultimately, a variety of cellular metabolites.

Transcriptome. The complete set of all RNA transcripts, including non-coding RNA transcripts characteristic of a given organism, or a specific set of transcripts (RNA molecules) present in cells of a particular type is called a **transcriptome**. Unlike the genome, which is usually the same for all cells of the same lineage, the transcriptome can vary greatly depending on environmental conditions. Since the concept of transcriptome includes all transcripts of a given cell, it also reflects the profile of gene expression at a given time, including maybe a manifestation of pathological processes (e.g., the expression of genes typical of cancer cells). One method for studying the transcriptome is single-cell RNA sequencing (RNA-Seq). This is the sequencing of RNA derived from a single cell, made possible by the development of next-generation high-speed sequencing (NGS) methods, which will be described in detail in Chapter 7. The use of bioinformatic methods makes it possible to assemble complete sequences of all transcripts and construct expression profiles of specific genes. The science of transcriptomes is called transcriptomics.

Proteome. The entire set of proteins expressed in a given cell type or an organism, at a given time under given conditions, is called a **proteome**. The science of the proteome is called **proteomics**. It deals with the extensive study of proteins, their structure, physiological role, and functions. The international Human Proteome Project is an extension of the Human Genome Project. In terms of the scale of the problem posed and the anticipated workload, the Human Proteome Project is much larger. While the human genome consists of about 20 thousand genes, the number of proteins in the human body due to the presence of various protein proteoforms exceeds 2 million. Moreover, unlike the genome, the proteome changes continuously depending on the influence of intracellular and external factors and is essentially a set of proteins fixed in time in a particular biological object and a particular situation.

Metabolome. The **metabolome** is a qualitative and quantitative set of all low-molecular-weight biomolecules (metabolites) present in the cell, which participate in various metabolic reactions and are necessary for the maintenance, growth, and normal functioning of the cell. The metabolome includes various compounds, including amino acids, fatty acids, carbohydrates, vitamins, lipids, etc. The number of different molecules in the metabolome varies depending on the organism being studied. It is also continually changing due to the various chemical reactions constantly taking place in the cell. All the metabolites in the metabolome form a large network of metabolic reactions, where the outputs of one enzymatic chemical reaction are inputs for others. For this reason, the metabolome is time-sensitive and much more dynamic than the proteome and genome. Metabolites, as products of gene expression and all other biological processes, have a high degree of diversity in their chemical properties.

CHAPTER 7

GENOMICS. TECHNIQUES OF MOLECULAR GENETICS

NUCLEIC ACID EXTRACTION METHODS

DNA and RNA extraction is an important step in preparation of samples for analysis and diagnostic procedures.

Molecular genetic analysis allows the study of virtually any biological material: blood, saliva, semen, hair, bone, muscle tissue, urine, amniotic fluid, chorionic villi, aborted fetal material, cadaveric material, nails, and more.

Stages of DNA (RNA) extraction:

1. Cell lysis (cell disruption), which destroys cellular structures and produces a lysate (cell-free extract).
2. Inactivation of cellular nucleases (DNases and RNases), which can break down the isolated molecules.
3. Extraction of the desired nucleic acid from the lysate.

There are 3 most common approaches used to extract nucleic acids.

Express methods are based on temperature lysis of cells and subsequent centrifugation. As a result, insoluble components precipitate at the bottom of the test tube, and the supernatant which contains DNA is used for analysis (Fig. 69). The method is easy to use, takes no more than 10–15 minutes, and provides maximum DNA yield, but is often inefficient. This method also cannot isolate RNA because the enzymes that degrade it (RNase) are resistant to heat.

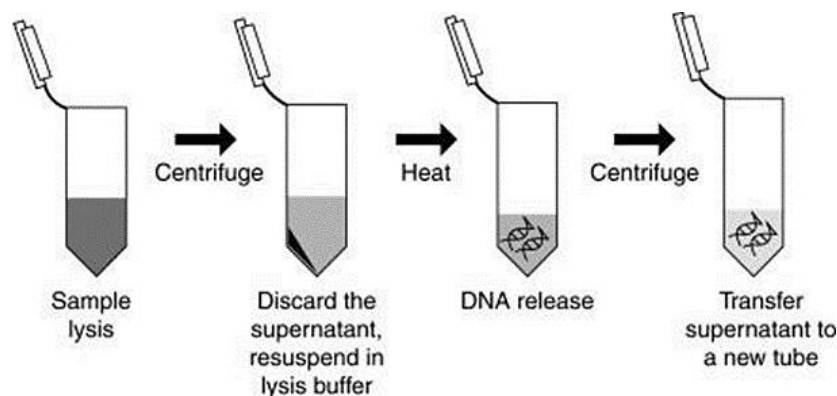


Figure 69. Boiling method of DNA extraction

Phenol-chloroform extraction. The cell lysate is prepared, phenol, chloroform, and isoamyl alcohol are added. After subsequent centrifugation, a solution with two phases is formed: organic (lower) phase with all lipids, and aqueous (upper) phase with nucleic acids. Proteins are situated between the phases. The DNA taken from the upper phase can be precipitated by adding ethanol and isopropanol. The precipitated DNA is separated by centrifugation. In the final step, the nucleic acid is dissolved in an aqueous buffer.

Sorbent isolation methods — differential sorption (sorption — absorption by a solid from the environment) of nucleic acids (DNA and RNA) on a solid carrier (most commonly: silica gel or glass beads). The nucleic acid binds to glass in the presence of high concentrations of specific salts. Other components of the clinical material remain in the solution. Using centrifugation, the nucleic acid carrier is precipitated and the supernatant is removed. A series of subsequent washes produce a highly purified nucleic acid preparation (Fig. 70).

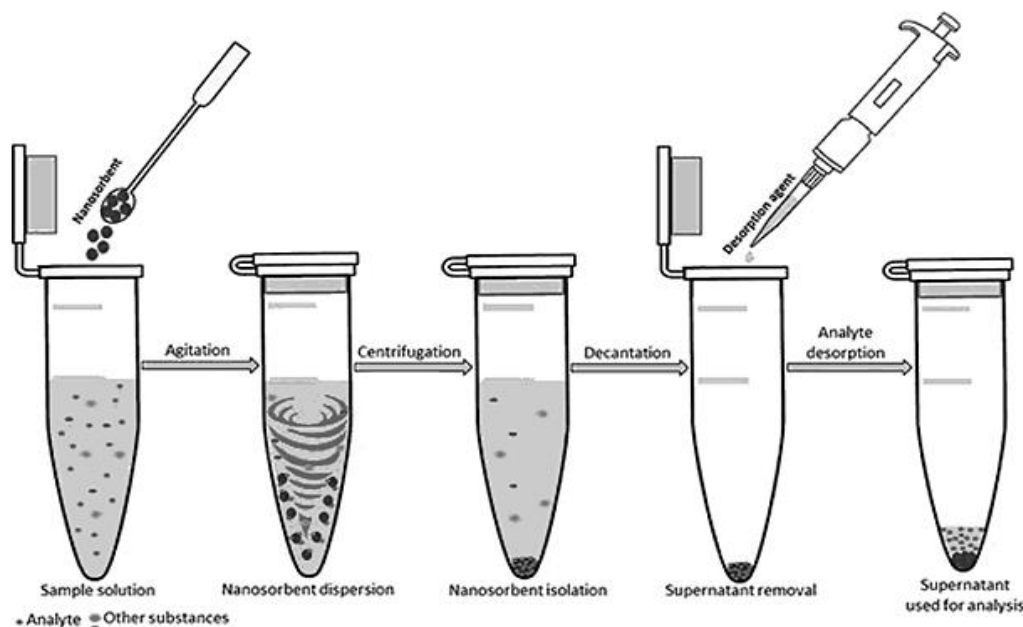


Figure 70. Sorbent isolation method of DNA extraction

ELECTROPHORESIS

Electrophoresis is a method that helps to separate DNA fragments of different lengths from each other. DNA molecules carry a negative charge, so they can be separated by electrophoresis in **agarose** or **polyacrylamide gel**. By comparing the mobility of the studied DNA with the mobility of DNA markers of known size, it is possible to determine the size of the DNA fragment. To visualize the results, DNA molecules are stained with **ethidium bromide** (EtBr).

Stages of the method:

1. DNA sample is cut into smaller fragments with restriction enzymes.
2. The fragments are placed into wells in a gel.
3. Electric current is applied to the gel.
4. As DNA has negative charge, DNA fragments start to move from the wells through the agarose gel in the direction from the negative electrode to the positive electrode.
5. Shorter DNA fragments move through the gel faster, so all the DNA fractions ultimately separate into fractions of different lengths.

RESTRICTION ANALYSIS

DNA restriction endonucleases (restriction enzymes, restrictases) are the enzymes that can cut DNA in specific sites (Fig. 71).

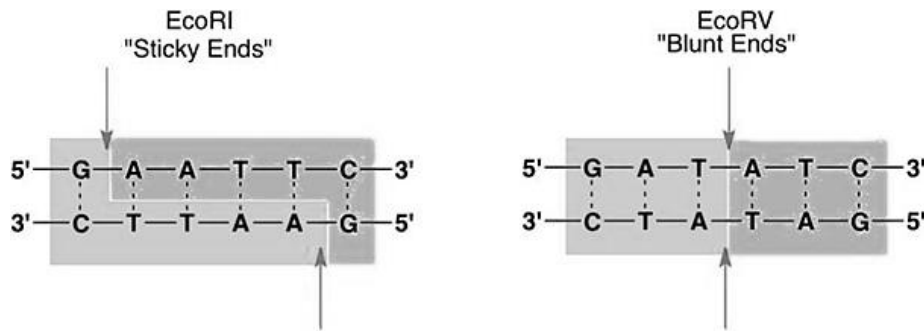


Figure 71. Cutting the DNA molecule with a restriction enzyme to form sticky and blunt ends

For example, the enzyme EcoR I recognizes the sequence $5' \dots \text{GAATTC} \dots 3'$ and cuts it between G and A in each strand. Notice that both strands contain the same sequence GAATTC, but one of them must be read in the opposite direction. The DNA containing such region is cut into two forming fragments $5' \dots \text{G}$ and $5' \text{AATTC} \dots 3'$ with ends $3' \dots \text{CTTAA} 5'$ and $\text{G} \dots 5'$. Both fragments contain identical single-strand regions TTAAG which can bind to each other (due to complementarity) forming a weak connection with hydrogen bonds. As such complementary single-strand regions can “stick” to one another, they are called sticky ends.

The nucleotide sequence recognized and cut by a restriction endonuclease (e.g. GAATTC for EcoR I) is called the *restriction site*. Restriction endonucleases cannot cut the DNA which does not contain proper restriction sites.

Restriction analysis is a method of double-stranded DNA analysis based on its processing with restriction enzymes and subsequent separation of the obtained fragments by electrophoresis. Since each restrictase cuts only certain sequence in DNA, processing the samples containing identical DNA with the same restriction enzymes produces same sets of fragments. Such set is unique for every species or strain. The diagram of DNA where restriction sites are marked is called a restriction map.

The “pattern” of restriction fragments obtained as a result of restriction analysis can be used for various purposes:

- for construction of restriction maps;
- to compare DNA from different people (DNA fingerprinting);
- to detect the presence of mutations (see restriction fragment length polymorphism);
- to detect a certain sequence with a DNA probe (see nucleic acid hybridization).

NUCLEIC ACID HYBRIDIZATION

Nucleic acid hybridization is the joining of two complementary nucleic acid chains from different sources to form double-stranded molecules. This process helps to identify specific DNA sequences. There are several types of nucleic acid hybridization techniques (Southern blotting and Northern blotting).

Southern blotting is used to identify specific DNA sequences, while Northern blotting is used to identify RNA sequences. Both methods use DNA probes. DNA probe is a short nucleotide sequence that is complementary to the target nucleotide sequence. Probes are labeled with a radioactive isotope or fluorescent dye.

In Southern blotting DNA fragments are separated into fractions with gel electrophoresis, and then they are transferred onto nitrocellulose or nylon membranes, where for hybridization with the probe occurs (Fig. 72). The method is named after its inventor, the English biologist Edwin Southern.

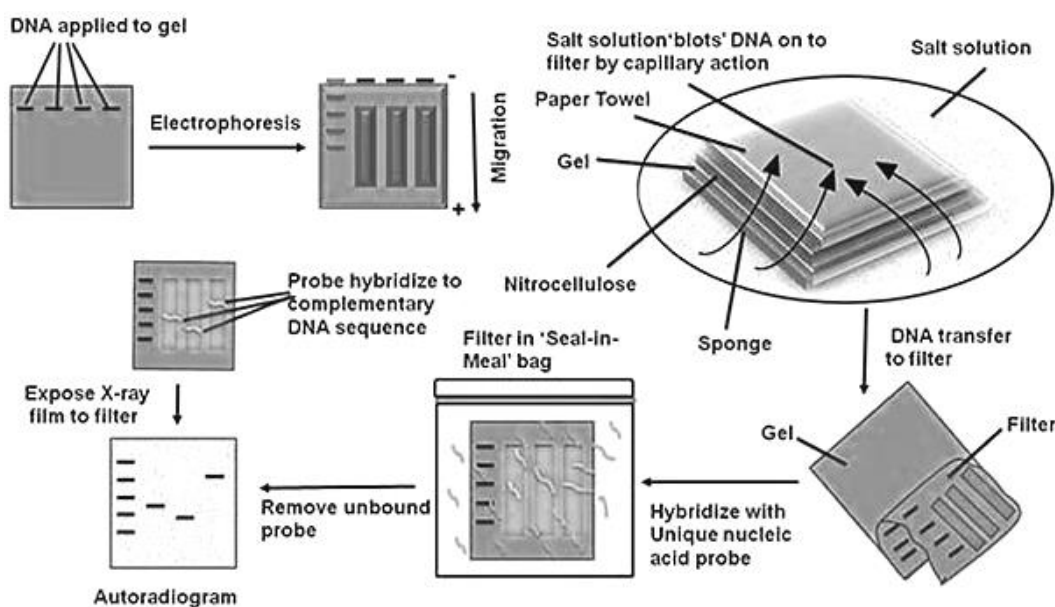


Figure 72. Southern blot hybridization

Stages of the method:

1. Restriction endonucleases cut DNA from the sample into smaller fragments which are then sorted by *gel electrophoresis*.
2. The gel with DNA is placed into an alkaline solution to denature the double-stranded DNA.
3. A nitrocellulose membrane is placed on the top of the gel. A stack of paper towels and weights are placed on top. A buffer solution moves by capillary action through the gel, membranes, and paper towels. It also takes the DNA to the membrane.
4. The membrane is processed and DNA permanently attaches to it.

5. The membrane is then exposed to a hybridization probe. The probe is a single-strand DNA fragment with a specific nucleotide sequence complementary to the target which must be detected. The probe is labeled with radioactive tags.

6. After hybridization, the membrane is washed to remove unhybridized probes.

7. The pattern of hybridization is visualized on X-ray film by autoradiography. Only the bands corresponding to the DNA fractions containing the target DNA are visualized.

DNA MICROARRAYS

DNA microarrays (DNA chips) allow to hybridize many DNA probes with their target DNA at the same time. A microarray is a microscope slide with thousands of fields, each containing unique single-stranded DNA sequences that serve as probes for the DNA fragments being sought. After hybridization, computer scans the slide and compares it with a map.

The main applications of the probes include:

- Transcriptome studies — analysis of all mRNA created by the cell reflects the profile of proteins created in that cell. The studies of cell transcriptomes can help to explain mechanisms making cells with same genotype different from one another.

- Studies of DNA variation: detection of mutations and single nucleotide polymorphisms (SNP).

- Detection of deletions and amplifications in the genome.

In this chapter the work of microarray is considered on the example of transcriptome analysis (Fig. 73).

1. First, mRNA is isolated from cells or tissues and subjected to reverse transcription to synthesize cDNA labeled with fluorescent dyes. If gene expression in different cells or tissue samples should be compared, cDNA from one tissue is labeled with one color (e.g., red) and that from the other one — with another dye (e.g., green).

2. The labeled cDNAs are denatured and incubated overnight with a microarray so that they hybridize with complementary DNA sequences.

3. The microarray is then washed and scanned with a laser, which causes the cDNA to fluoresce.

4. Because the locations of the different probe fields are known, the pattern of fluorescent spots shows which genes are expressed in the tissue of interest, and the intensity of the fluorescence reflects the relative level of expression. The brighter the stain, the more of the mRNA is expressed in that tissue.

The biggest advantage of DNA microarrays is that they allow thousands of genes to be studied simultaneously.

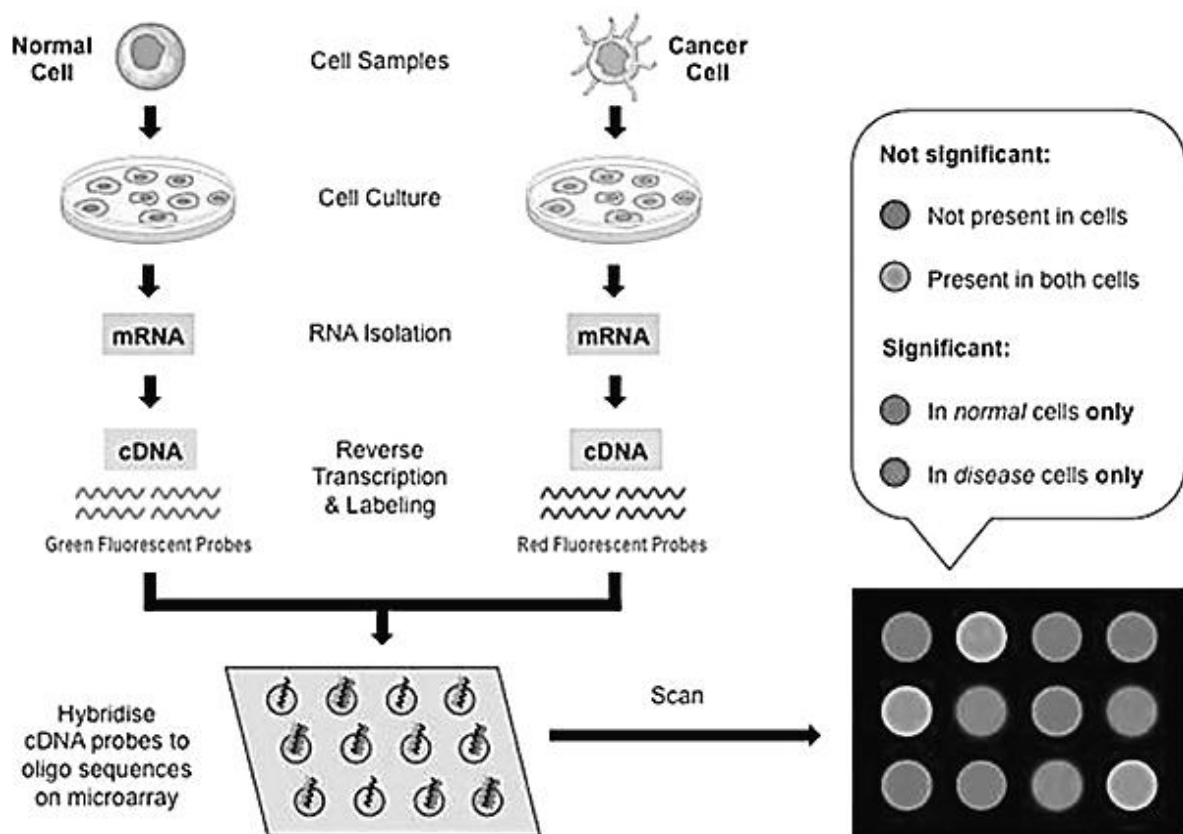


Figure 73. The principle of DNA microarrays

POLYMERASE CHAIN REACTION (PCR)

Polymerase chain reaction (PCR) is the technique used for amplification of desired DNA sequence. DNA amplification is the production of multiple copies of the DNA.

Standard PCR. PCR uses DNA-polymerase to copy DNA from the sampl. The reaction occurs in a tube that is placed into a thermocycler. Thermocycler is a device that provides periodic heating and cooling of test tubes. The tube contains the following components (Fig. 74):

1. **Analyzed DNA**, which serves as a template for multiple copying of the desired fragment.

2. **Primers** — short (20–30 nucleotide long) DNA fragments from which DNA-polymerase can start making the new strand. They are complementary to a small region of the DNA sample. There are two types of primers, because the DNA template consists of two strands having different sequences. Since DNA polymerase can add new nucleotides only to the 3' end of the primer, it moves in only one direction. This property of the enzyme is used to copy only the region of interest: the primers bind to different strands on the two sides of the region.

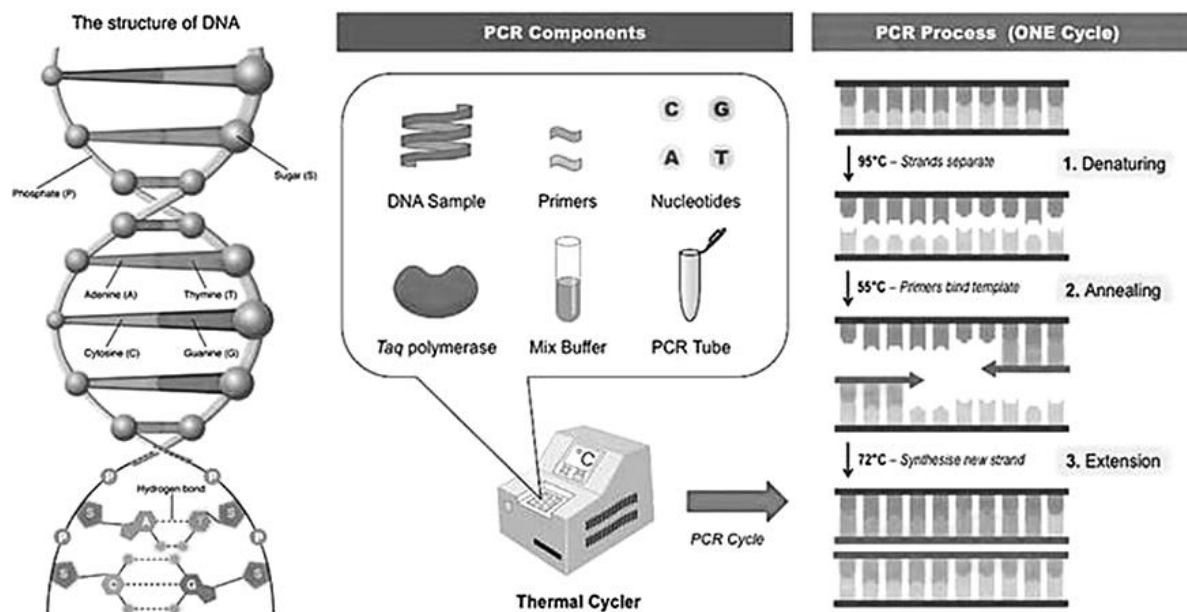


Figure 74. PCR

3. **Nucleotide triphosphates:** dATP, dTTP, dCTP, and dGTP. DNA polymerase includes nucleotide monophosphate into the new strand, two other phosphates are released in the form of P_2O_7 .

4. **DNA polymerase** — the enzyme that adds complementary nucleotides to the template and makes strand. It can extend the 3' end of the primer. Polymerase chain reaction requires periodic heating to temperatures at which proteins denature. That is why common polymerase used in PCR is a thermostable enzyme isolated from thermophilic bacteria or archaea, e.g. Taq-polymerase from *Thermus aquaticus*.

5. **Buffer** — the solution containing various ions to maintain the correct pH, magnesium salts necessary for the polymerase to work, and other components necessary to prevent the reaction components from sticking to the walls of the test tube.

PCR consists of three steps:

1. **Denaturing.** At this stage the two strands of DNA separate from each other. The hydrogen bonds between them are broken by temperature: thermocycler heats the tube to 94–98 °C. This is the shortest stage of the reaction; it may last about 15 seconds.

2. **Annealing** of the primers. Temperature is lowered to 50–65 °C to allow primers to bind with complementary regions of the DNA strands.

3. **Extension** (or DNA synthesis). This step is performed at 72 °C, the optimal temperature for Taq polymerase. The enzymes attach to primers and add complementary nucleotides to the templates extending the 3'-end of each primer. This is the longest stage of PCR which may take several minutes.

These 3 steps are repeated 20–40 times, and each cycle doubles the amount of the desired DNA in the tube. Each newly synthesized DNA strand

becomes the template in the next cycle. Thus, the amount of the desired product increases exponentially during the reaction.

The result of PCR can be estimated by electrophoresis with subsequent visualization.

REVERSE TRANSCRIPTION PCR (RT-PCR)

This type of PCR is used when RNA, but not DNA is available. It helps to study RNA viruses or mRNA (to determine the rate of gene expression). Since the polymerase chain reaction uses DNA as a template, reverse transcription is used first: the enzyme reverse transcriptase builds complementary DNA (cDNA) on the RNA template. Then standard PCR is performed with the cDNA.

The procedure can be performed in two tubes. In the first one, cDNA is produced from the RNA at 37°C. Then standard PCR is performed in the second tube. It is possible to use one tube: all reagents are added, and left for 1 hour at 37 °C. After that, the tube is placed in a thermocycler. RT-PCR can also be performed with a single enzyme — Tth polymerase. This thermostable polymerase was isolated in 1985 from the bacterium *Thermus thermophilus*.

The enzyme has dual activity: in the presence of magnesium ions, it has polymerase activity, and in the presence of manganese ions, it has reverse transcriptase activity. Both reactions can proceed at 70 °C.

Real-time PCR (quantitative PCR, qPCR) and RT-qPCR. In real-time PCR the amount of target DNA in the tube is constantly measured. That is why the result of PCR can be known even when the reaction is not yet over. This also allows to calculate the initial number of DNA in the sample. For this reason, this type of PCR is also called quantitative PCR, or qPCR. This technique can also use RNA as a template. In this case, it should be started from reverse transcription.

Real-time PCR uses DNA probes labeled with **fluorophores** — molecules that can glow in specific conditions. The other molecule attached to the probe prevents fluorescence (**quencher**). The probes bind to DNA, so they are on the way of DNA polymerase. The enzyme breaks them when builds the new strand. When the probe is destroyed, the fluorophore separates from quencher and starts to fluoresce. With every PCR cycle the amount of free fluorophore increases, and this amount is measured and used for calculation of DNA molecules in the sample (Fig. 76).

Intercalating dyes (Latin: intercalatio — insertion, addition) are molecules capable of reversibly embedding into double-stranded DNA or RNA (Fig. 76). The intercalating dye starts to fluoresce only when it binds with DNA, and there is no fluorescence in absence of DNA. With every cycle the amounts of DNA increase, and this, in turn, causes increase in fluorescence. The intensity of fluorescence helps to calculate the amount of DNA in the sample.

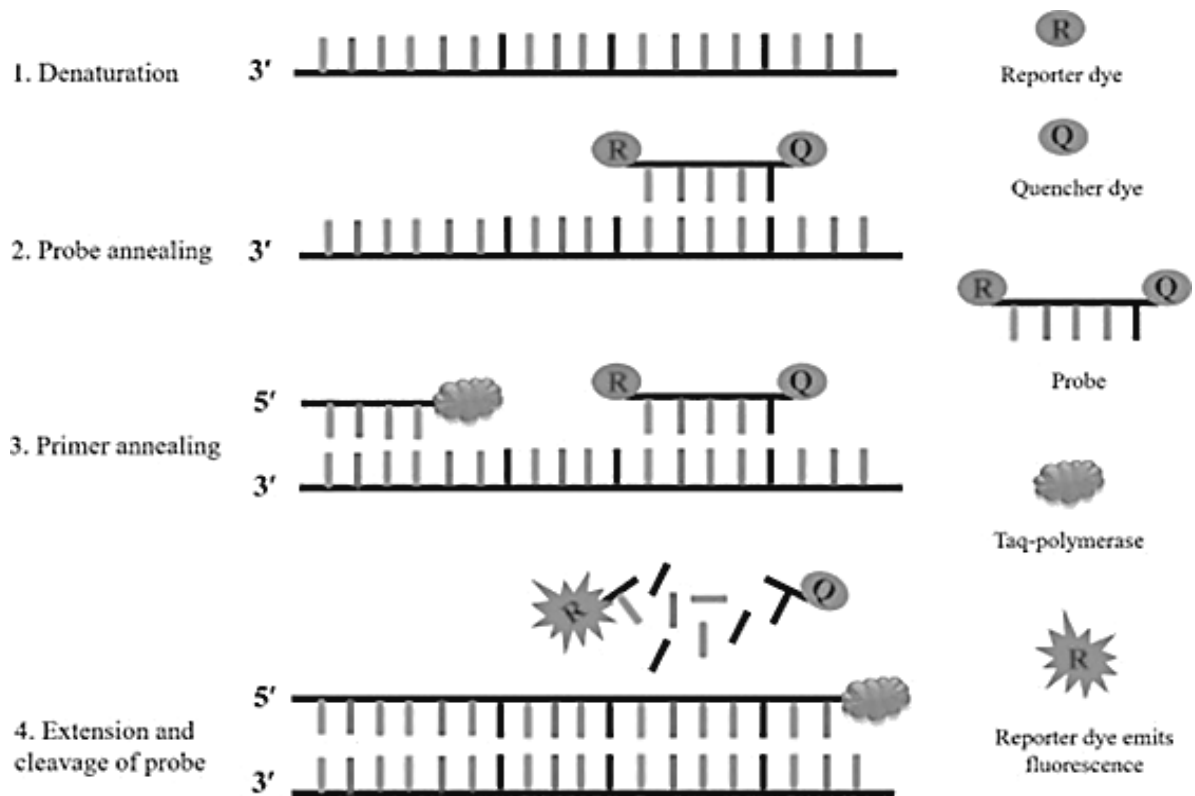


Figure 75. The fluorophore probe

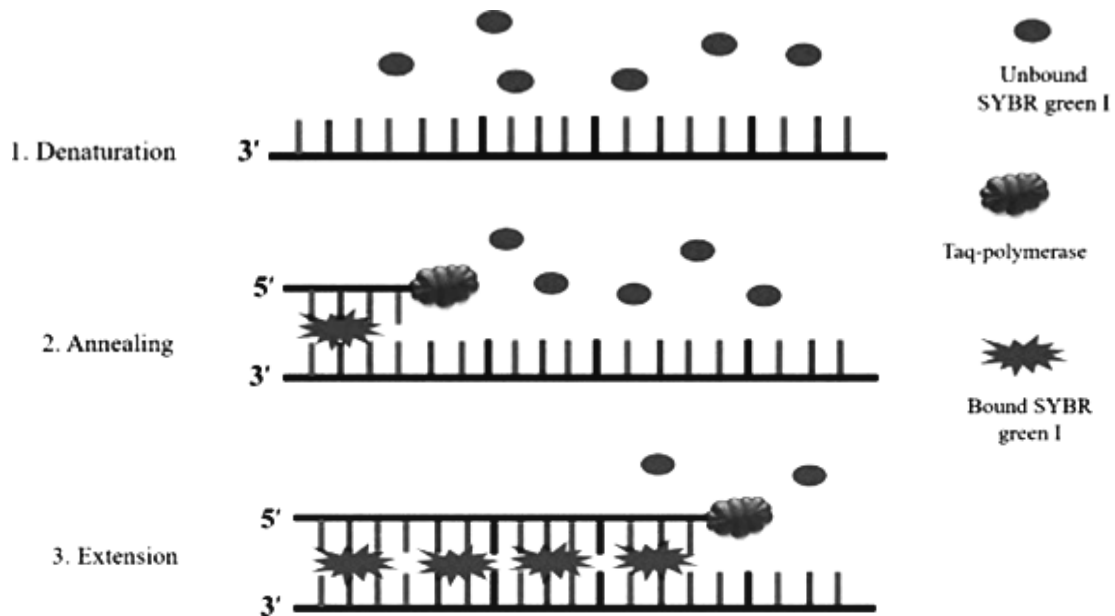


Figure 76. The principle of action of the intercalating dye

MULTIPLEX PCR

This variant of PCR is used when several different DNA sequences need to be detected at the same sample. A set of primers is added into one tube, and several target fragments are amplified simultaneously.

The results can be evaluated by gel electrophoresis (Fig. 77). Each set of primers should produce a fragment that is different in size from others. Otherwise, the bands with different DNA in the gel will overlap.

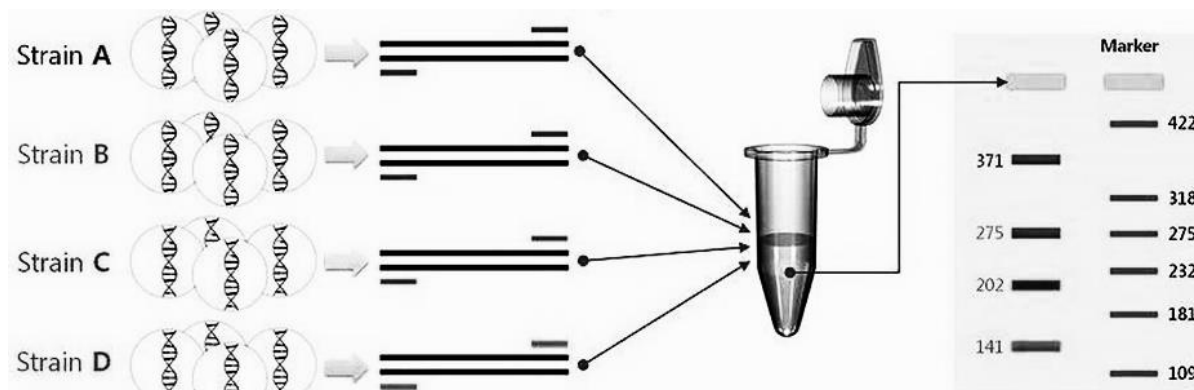


Figure 77. Multiplex PCR scheme (4 primer variants used)

DNA SEQUENCING

The **sequencing** of nucleic acids is the determination of their nucleotide sequence. The development of DNA sequencing technologies began in the 1970s. Since then, three generations of these technologies with fundamentally different methodological approaches arose. Examples of such techniques are: Sanger sequencing, pyrosequencing, nanopore sequencing, bisulfite sequencing.

Sanger sequencing was developed in 1977 by F. Sanger and his colleagues. The principle of this procedure was to create a large number of fragments of sequenced DNA whose length differed by 1 nucleotide. For example, to “read” the “TCAGAG” sequence, we first synthesize fragments containing **T**, **TC**, **TCA**, **TCAG**, **TCAGA**, and **TCAGAG** at the end of the sequence. Due to some peculiarities of the procedure, the last nucleotide in each chain is always known. Then, due to the different lengths of the fragments, they can be sorted via gel electrophoresis. The sequence of fractions in the gel demonstrates the nucleotide sequence of the original DNA.

Nowadays, the Sanger sequencing procedure is fully automated and is performed by devices called sequencers:

1. The first step of sequencing resembles PCR: DNA polymerase, four types of deoxynucleotides (dNTP), primers, and samples of sequenced DNA are used.

2. DNA polymerase, which synthesizes a new DNA strand, adds a new nucleotide to the 3'-OH group of the previous one. If the nucleotide does not have a 3'-OH group, further chain synthesis will stop there. This principle is used to create DNA fragments of different lengths: a certain proportion of dideoxynucleotides (ddNTP) without a 3'-OH group is added to the reaction mixture, so DNA polymerase always has a small chance to add

a dideoxynucleotide to the growing DNA chain instead of the usual one (Fig. 78). Because of this, the Sanger method is also called the dideoxy method or the chain termination method.

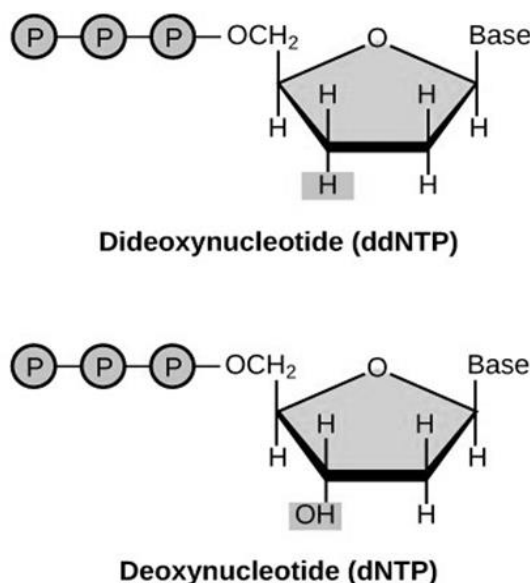


Figure 78. Deoxynucleotide triphosphate (reactive OH- group at the 3' carbon atom) and dideoxynucleotide triphosphate (no reactive OH- group at the 3' C atom)

3. As mentioned above, the fragments are separated from each and sorted by their length via gel electrophoresis. Dideoxynucleotides with A, T, G, and C are initially labeled with fluorescent dyes of different colors. Thus, the DNA fractions containing fragments ending with different nucleotides can be distinguished from each other. Hence, the order of colored DNA fractions in the gel represents the nucleotide sequence of the DNA sample.

Gel electrophoresis is performed in a thin capillary. During this process, a laser activates the dyes, and each of the four types of dideoxynucleotides emits a different color, which is recorded by a photodetector.

Sanger sequencing allows to “read” the fragments ranging in length from a few hundred to 1,000 nucleotides.

PYROSEQUENCING

Pyrosequencing (belongs to the second-generation methods) is the sequencing of nucleotides in DNA using the “sequencing by synthesis” principle.

When DNA-polymerase adds a complementary nucleotide into the new strand, it releases pyrophosphate (P_2O_7) as a byproduct. Presence of pyrophosphate can be detected: due to enzymes sulfophosphate (APS) and ATP-sulfurylase it is used to produce ATP. This molecule, in turn, is a source of energy for another reaction — oxidation of luciferin to oxyluciferin with emission of a quantum of light. The intensity of the light released

(chemiluminescence) is recorded (Fig. 79). Thus, a small flash of light means that current nucleotide is added into a polynucleotide chain.

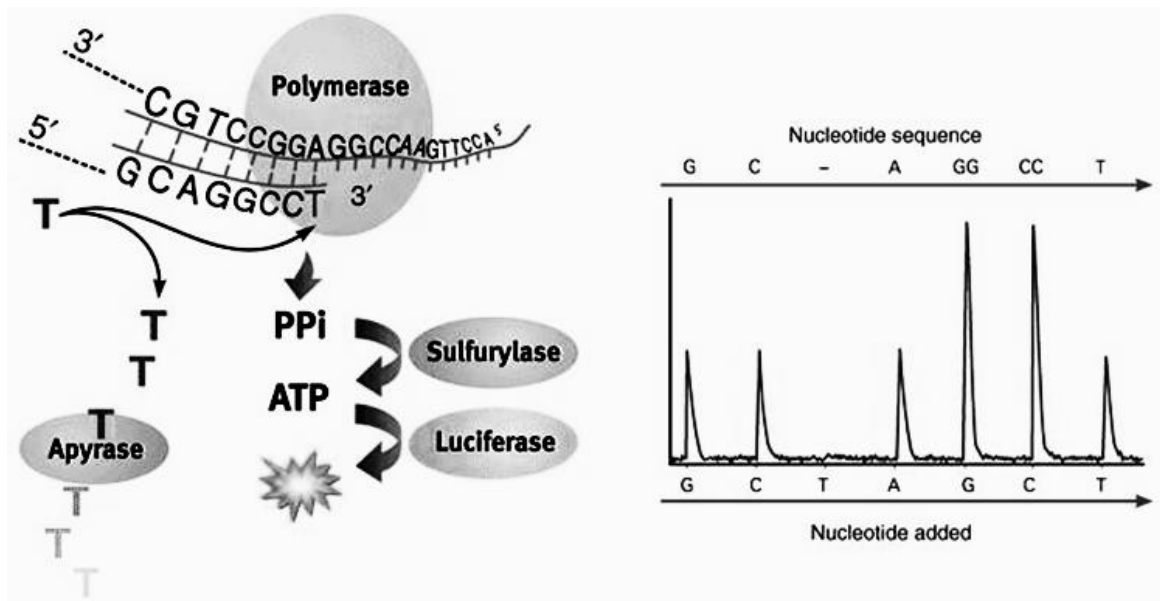


Figure 79. Pyrosequencing principle

The analysis includes the following stages:

1. DNA is cut into fragments 300–500 bp long. They are bound with primers and adapters and denatured into single-stranded molecules.

2. Emulsion PCR: DNA fragments, PCR reagents, and special beads are mixed in an aqueous solution. A water-in-oil emulsion consisting of tiny droplets is formed (Fig. 80). Each bead is bound with multiple copies of the same sequence.

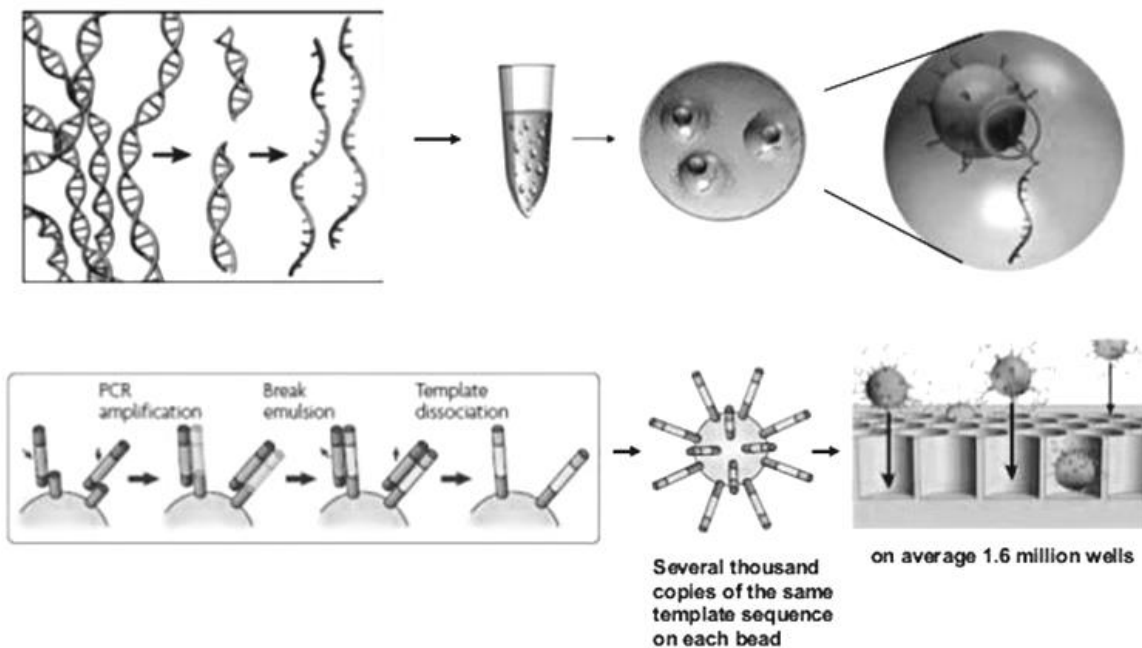


Figure 80. Emulsion PCR

3. Sequencing is performed on a plate with many small chambers. One chamber has only one bead (and only one sequence in all DNA molecules bound to it), along with enzymes and substrate that generate the chemiluminescent reaction signal.

4. Nucleotide solutions (A, T, G, C) are subsequently passed through the plate. Every time the previous nucleotide is washed off the plate. If current nucleotide is included into new strand, a light signal is produced. Absence of the signal from the chamber means that the nucleotide doesn't match current nucleotide of the template strand on the bead (Fig. 81).

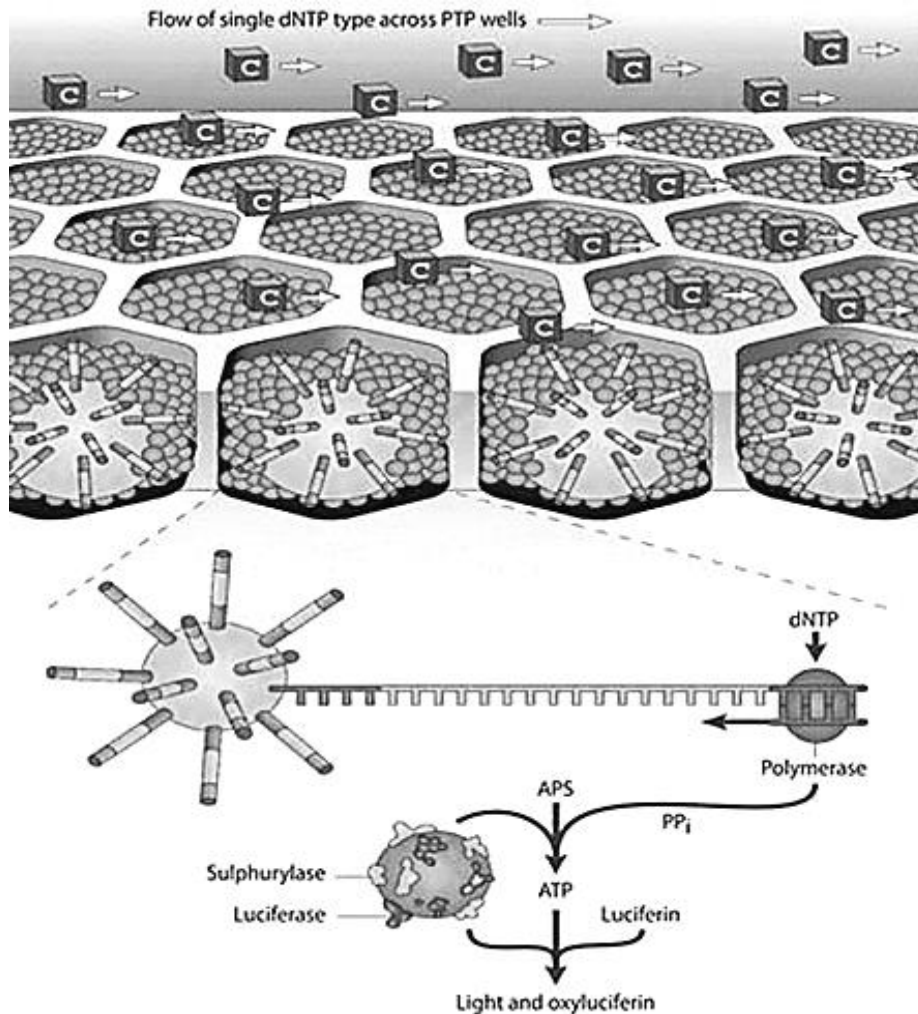


Figure 81. Pyrosequencing scheme

5. The result of sequence analysis in an individual chamber looks like a sequence of bars whose color corresponds to one of the T, C, A, or G nucleotides.

6. The last step of sequencing is the assembly of the fragments into the original DNA sequence.

NANOPORE SEQUENCING

Nanopore sequencing is one of the third-generation methods.

The method is based on measuring ion current through a single nanopore in a non-conductive membrane. As nucleotides pass through this pore, the current intensity drops. The time by which the ion current changes and the magnitude of this drop depend on which nucleotide is currently inside the pore.

A nanopore system is a reaction chamber with an electrolyte solution inside. The chamber is divided into two parts by a lipid membrane in which a single nanopore is embedded. A voltage is applied to the parts of the chamber, causing ions to flow through the pore. When the molecules under study pass through the pore in the direction of the field, they decrease the cross-section available to the ions, and the current intensity drops. By analyzing the change in the current, we can determine the properties of the molecule passing through the pore.

In the case of nucleic acids, the nanopores used are several nanometers in diameter, so that DNA and RNA can only pass through the pore in single-stranded form, but not in double-stranded form. As the nucleic acid molecule passes through the pore, individual nucleotides are trapped at specific sites within the pore, resulting in a measurable drop in current strength. By reading the change in current, the nanopore sequencer determines the nucleotide passing through the pore at a specific time interval (Fig. 82).

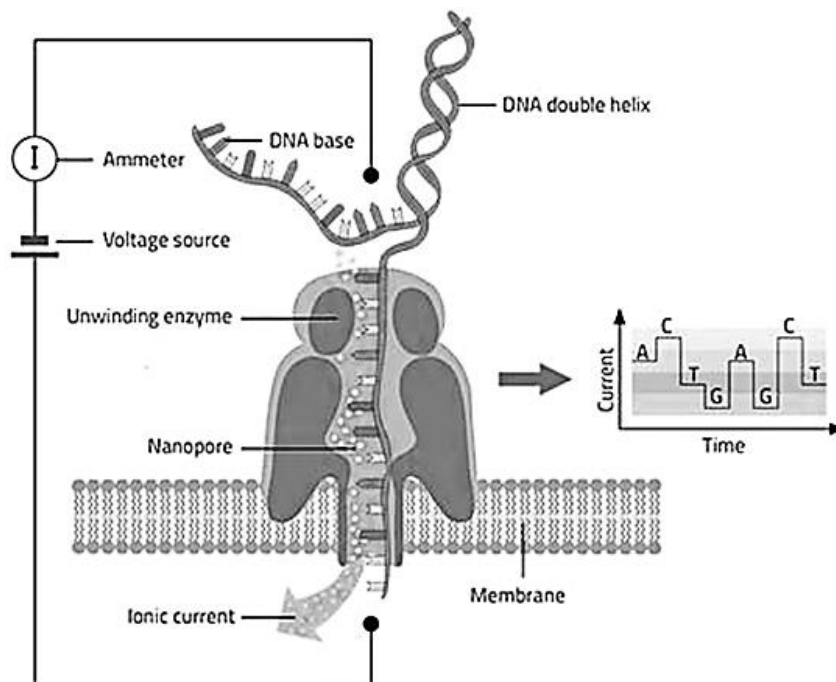


Figure 82. Nanopore sequencing technique

Nanopore sequencing technology allows direct reading of DNA or RNA strands online, and the portability of the equipment enables sequencing even in the field.

BISULFITE SEQUENCING

Bisulfite sequencing is the general name of a group of methods aimed at studying DNA methylation by treating it with bisulfite. Recall that methylation is epigenetic labeling. It affects the level of gene expression by inhibiting transcriptional activity.

Bisulfite acts on single-stranded DNA, converting cytosine to uracil. If this cytosine is methylated, i.e., a methyl group is attached to its fifth carbon atom, such cytosine is not transformed. Thus, bisulfite changes the DNA sequence depending on its methylation pattern, and after exposure to it, one can determine which CpG dinucleotides have been methylated by comparing the changed sequence with the original one.

Stages of bisulfite sequencing:

1. First, genomic DNA is cleaved with a methylation-insensitive restriction enzyme. This allows the enzyme to cleave both methylated and unmethylated regions. Typically, MspI is used. This enzyme targets the 5'CCGG3' sequences and cleaves the phosphodiester bonds upstream of the CpG dinucleotide. Using this particular enzyme, each fragment has a CpG at each end.

2. The DNA fragments are then treated with bisulfite, which results in the deamination of unmethylated cytosine into uracil. The methylated cytosines remain unchanged because of the methyl group that protects them from the reaction.

3. PCR amplification. Bisulfite-transformed DNA is then amplified using PCR with primers.

4. The fragments are then sequenced. Consistency with the reference genome allows programs to identify methylated base pairs in the genome.

CHAPTER 8

GENETIC ENGINEERING

GENETIC ENGINEERING: GOALS, OBJECTIVES AND STAGES

Genetic engineering is a set of techniques, methods, and technologies for the creation and use of recombinant RNA and DNA. **Recombinant DNA** is DNA created *in vitro* (i.e. under laboratory conditions rather than in a living organism) by combining two or more DNA fragments from different sources.

Genetic engineering aims to construct molecular recombinant DNA and introduce them into a living organism. As a result, the organism acquires unique genetic, biochemical, and physiological properties. The process of introducing a gene into a living organism is called **transgenesis**.

The practical tasks of genetic engineering are to create genetically engineered bacterial strains for the production of drugs and vaccines; to create transgenic plants with desired properties; to create transgenic animals for practical purposes; to develop methods of human gene therapy.

The process of obtaining a particular sequence of DNA and creating multiple copies of it through replication in a living organism is called **DNA cloning (gene cloning)**.

Stages of genetic engineering:

1. Obtaining genetic material (DNA isolation).
2. Analysis of DNA fragments.
3. Construction of recombinant DNA able to replicate autonomously *in vivo* (in a living cell).
4. Transferring recombinant DNA into a host cell.
5. Selection of cell clones containing hybrid DNA molecules.
6. Use of modified cells.

Methods of obtaining genes for transgenesis:

1. Artificial gene synthesis. Artificial gene synthesis has been developed to obtain DNA fragments with known nucleotide sequences. The synthesis process is automated and carried out in DNA synthesizers.

Double-stranded DNA is synthesized in two steps. First, single-stranded oligonucleotides (i.e., short chains of nucleotides) are synthesized. Then they are linked by ligases and annealed (allowing double-stranded DNA molecules to form).

The principle of the method used for oligonucleotide synthesis:

- The first nucleotide is attached to a solid phase and placed in a reactor column.
- A second monomer and other chemical reagents are then passed through the column.
- The second monomer joins the first monomer to form a dimer.

- The column is then washed to remove unreacted reagents.
- These steps are repeated until the desired sequence is made.
- After synthesis is complete, the oligonucleotide is separated from the solid phase and purified (Fig. 83).

This technique is also used in the synthesis of oligonucleotides used as probes, primers, etc.

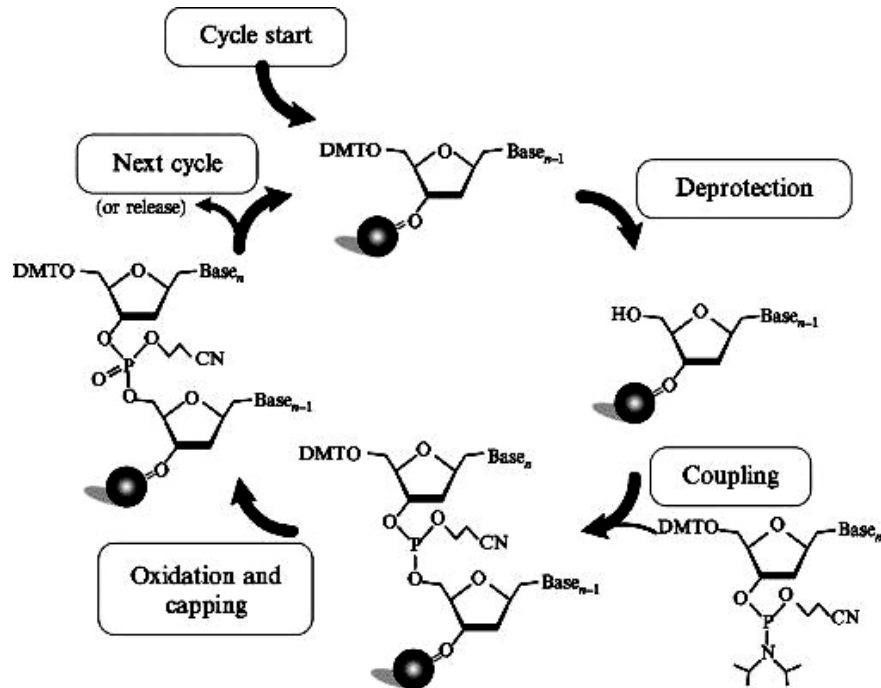


Figure 83. Artificial gene synthesis

2. Reverse transcription.

Reverse transcription is the synthesis of DNA that is complementary to mRNA (cDNA). This process is performed by **reverse transcriptase (revertase)**. Almost all eukaryotic mRNAs contain the poly(A) tail at their 3'-ends. This allows using oligo(dT) (a short chain of nucleotides with thymine) as a primer for reverse transcriptase. Reverse transcriptase extends the primer and produces a DNA strand. Then mRNA is destroyed with NaOH, and DNA polymerase is used to complete the synthesis of the second DNA strand. At the end of the synthesis, the first and

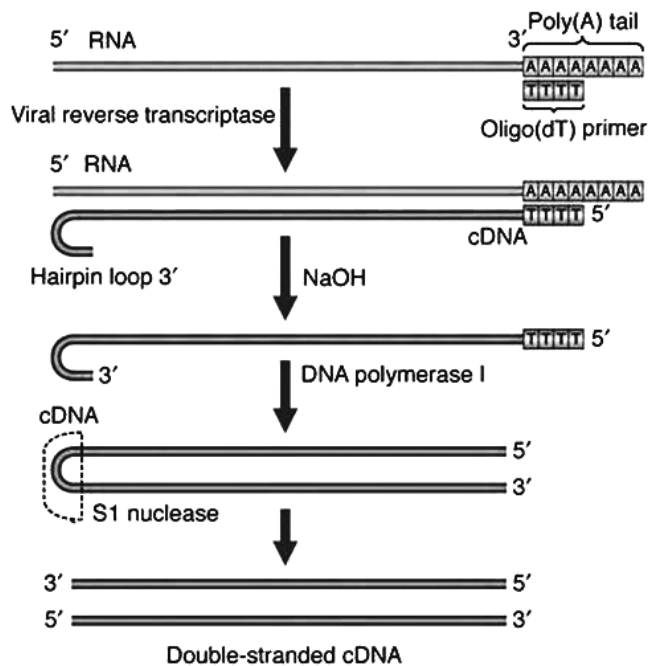


Figure 84. Scheme of cDNA synthesis on mRNA

second DNA strands are connected with a hairpin. It is cleaved with a special endonuclease, which can specifically hydrolyze single-stranded sections of nucleic acids (Fig. 84). The result is a complete DNA copy of the gene, which can be inserted into a vector and transferred into a host cell.

3. Natural gene isolation using restriction enzymes. The essence of this method is to “cut out” the desired fragment from DNA using **restriction enzymes** (or **restriction endonuclease**). Restriction endonucleases are enzymes that recognize specific sequences in DNA and cut them (see previous chapter). The restriction enzymes used in genetic engineering recognize *palindromic* sequences of double-stranded DNA and cut both the strands. A palindrome is the sequence which is identically read on different strands (from 5' to 3' end). For example: 5'GGCC3' is complementary to 3'CCGG5'.

The sequences recognized and cut by restriction enzymes are called **restriction sites** (Fig. 85). Some restriction enzymes cut DNA strands between nucleotides at opposite points to form **blunt ends**. If the cuts are some distance apart, the resulting ends contain single-stranded extensions that are complementary to each other (**sticky ends** are formed).

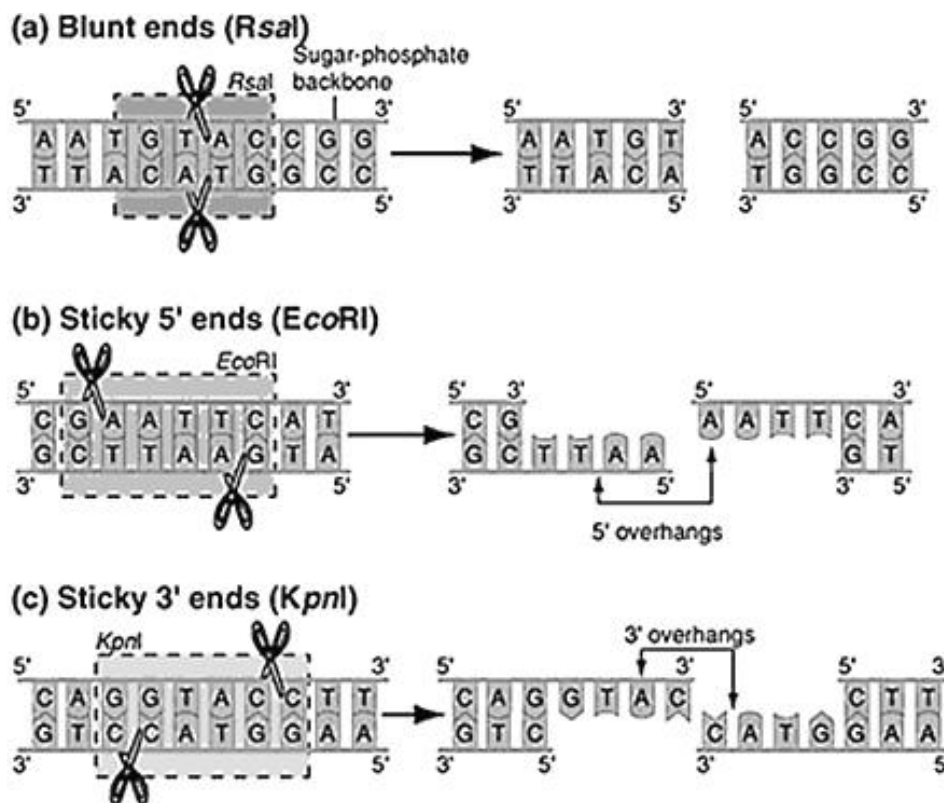


Figure 85. Diagram of DNA cutting by RsaI restrictases with the formation of blunt ends and EcoRI with the formation of sticky ends

More than 4,000 restriction enzymes capable of cutting DNA at more than 360 different sites have now been isolated and characterized. Gene isolation using restriction enzymes has some drawbacks:

- It is not always possible to select restriction enzymes that can cut precisely the section of DNA that contains the desired gene.
- The excised DNA fragment may contain introns, and recombinant DNA cannot be expressed in prokaryotic cells due to the lack of processing and splicing capabilities.

RECOMBINANT DNA

Currently, genetic engineering works with a variety of living systems (bacteria, yeast, plants, insect and mammalian cell cultures, whole living organisms). The choice of a system for a particular purpose is based on economic, ergonomic, and ethical considerations.

Commonly genetic engineering deals with a bacterium *Escherichia coli*, baker's yeast (*Saccharomyces cerevisiae*). *E. coli* is used in cloning and testing recombinant DNA, production of some biotechnological products. *S. cerevisiae* is most commonly used to study and express eukaryotic genes.

Whatever model organism is chosen, the gene itself is not functional in a host cell. For this reason, it should be inserted into another DNA — a vector. **Vector** is a DNA molecule that can carry inserted DNA and can be perpetuated in a host cell. The vector containing the insertion of the desired gene is called **recombinant DNA**.

Requirements for vectors:

- Presence of origins of replication (replication origin, ori) and replicate autonomously in the host cell.
- Stable inheritance by daughter cells.
- Presence of suitable restriction sites. Sometimes restriction sites are concentrated in the same small region of the vector. This region is called polylinker or MCS (multiple cloning site).
- Sufficient capacity to clone large genes.
- Presence of marker genes — the genes helping to recognize the cells having the desired gene and select them from other cells.
- The presence of the sequences ensures the efficient expression of the desired gene in the selected host if required (e.g. promoter, terminator, ribosome binding site).

MARKER GENES

Marker genes are the genes that give their hosts an easily detectable trait, which makes it possible to trace the hosts and estimate cloning efficiency. Among them, the following are distinguished:

1. Selectable marker genes are genes that make it possible to select their carriers in certain environmental conditions. Usually, they are the genes for the resistance to antibiotics, heavy metal salts, and herbicides (in plants).

For example:

- *amp* — ampicillin antibiotic resistance gene;
- *tet* — tetracycline antibiotic resistance gene;
- *cat* — chloramphenicol antibiotic resistance gene.

2. Reporter genes encode cell-neutral proteins, the presence of which is easy to visualize. For example:

- *lacZ* is the gene coding for beta-galactosidase — the enzyme that allows bacteria to produce blue dyes from another colorless molecule called X-gal;
- *gfp* — the gene coding the green fluorescent protein. Fluorescence appears when the transformed cells are illuminated with blue light.

Vectors are classified according to the hosts they are to enter (bacterial, yeast, etc.), according to their origin (plasmid vectors, viral vectors, shuttle vectors, artificial chromosomes), and according to their functional purpose (vectors for cloning, expression, sequencing).

We will consider vectors designed for DNA cloning (**cloning vectors**).

PLASMID VECTORS

In addition to the nucleoid, bacteria usually have small circular DNA molecules of several thousand base pairs long in the cytosol — plasmids. The number of plasmid copies in a cell can vary significantly. It depends on the genetic features of both the cell and the plasmid.

Typically, plasmids contain genes of resistance to antibiotics, heavy metal ions, or genes controlling the catabolism of certain organic compounds, etc. Some natural plasmids have been artificially modified and turned into cloning vectors. The capacity of plasmid vectors is small, ranging from 1–10 kb (1 kb = 1000 nucleotides).

One of the best-known and most used **plasmids**, **pBR322**, is an example of a cloning vector (Fig. 86). Its features are typical of plasmid vectors.

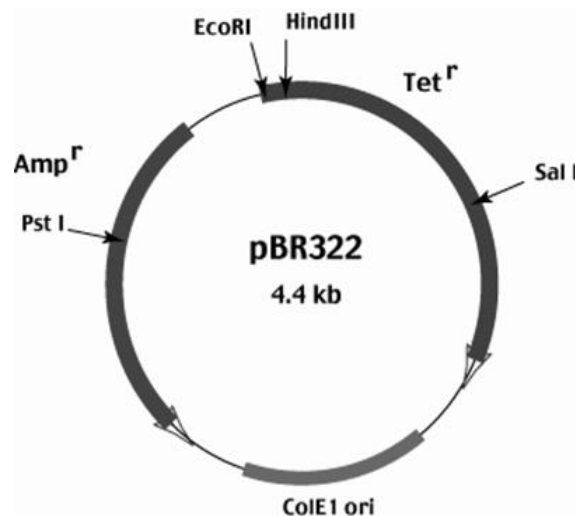


Figure 86. The plasmid pBR322

This plasmid is smaller in size than its natural analog, which makes it more resistant to damage and increases the efficiency of uptake by the bacterium during transformation.

The presence of replication origin ensures that the plasmid is replicated by the host cell. The plasmid pBR322 has ori allowing it to replicate more frequently than the bacterial chromosome. As a result, a large number of plasmid copies (100–200) are produced in the cell.

The pBR322 has genes providing the host with resistance to the antibiotics ampicillin and tetracycline. If cells are seeded on a medium with antibiotics, only those containing the plasmid will grow to form colonies.

There are restriction sites for various enzymes at various points in the pBR322. This allows cutting the plasmid and inserting desired DNA fragment into it. Then the plasmid with DNA insertion is closed back into a circular form. The sites for some restriction enzymes are located within the genes of antibiotic resistance. Therefore, the bacteria carrying the desired recombinant DNA can be detected due to the loss of antibiotic resistance.

In a series of **plasmids** called **pUC** (Fig. 87), the restriction sites are concentrated in one region called the polylinker that is situated within *LacZ α* . The gene allows bacteria to produce blue substance from colorless X-gal. However, if the gene is disrupted by the insertion of a foreign DNA fragment, the blue product is not synthesized. Bacterial cells containing the pUC recombinant plasmid can be easily detected by the absence of coloring if the substance X-gal is added to the medium.

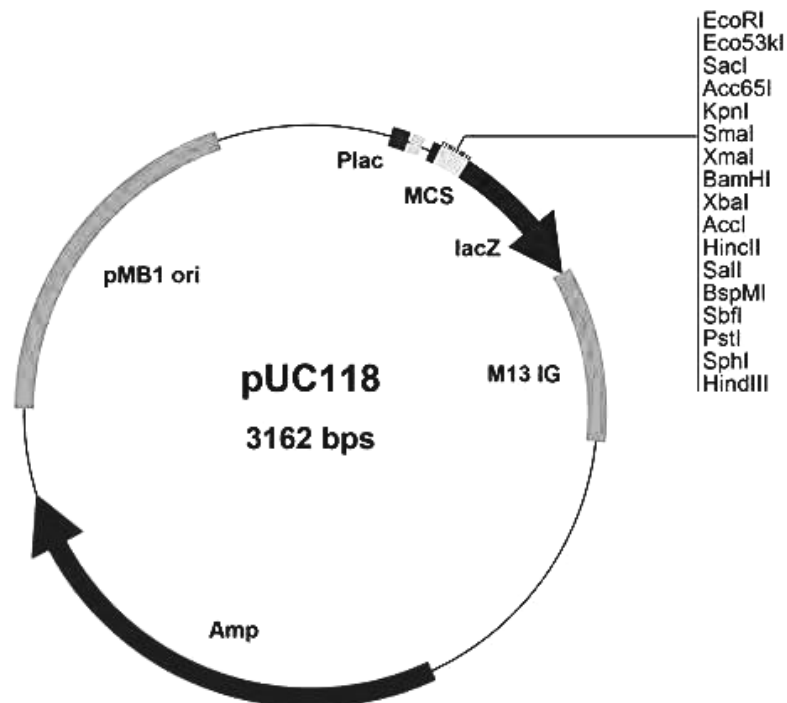


Figure 87. Plasmid from the pUC series

PHAGE (VIRAL) VECTORS

An important characteristic of a vector is its capacity. The insertion of foreign DNA over 5,000 bp significantly reduces the efficiency of vectors, so bacteriophages (phages, or bacterial viruses) are used as vectors to reproduce large fragments of DNA in bacterial cells. Cloning vectors based on phage λ (lambda) and phage M13 are most commonly used. The capacity of phage vectors is 5–25 kb.

Vectors based on phage λ (Fig. 88). Phage λ contains 48,500 bp double-stranded DNA in the form of a linear molecule with 12 bp long single-stranded cos-sequences (cohesive sites) that are complementary to each other. These sequences are like sticky ends and allow the phage DNA to close into a ring. If the large central region responsible for the lysogenic cycle (not essential for cloning) is removed from the phage DNA, a foreign gene up to 25 kb in size can be inserted instead. The resulting recombinant DNA assembles into protein viral particles *in vitro*. Bacterial cells are infected with the phage. There it multiplies, causes cell lysis, escapes into the medium, and infects other bacterial cells. The recombinant DNA is isolated and used.

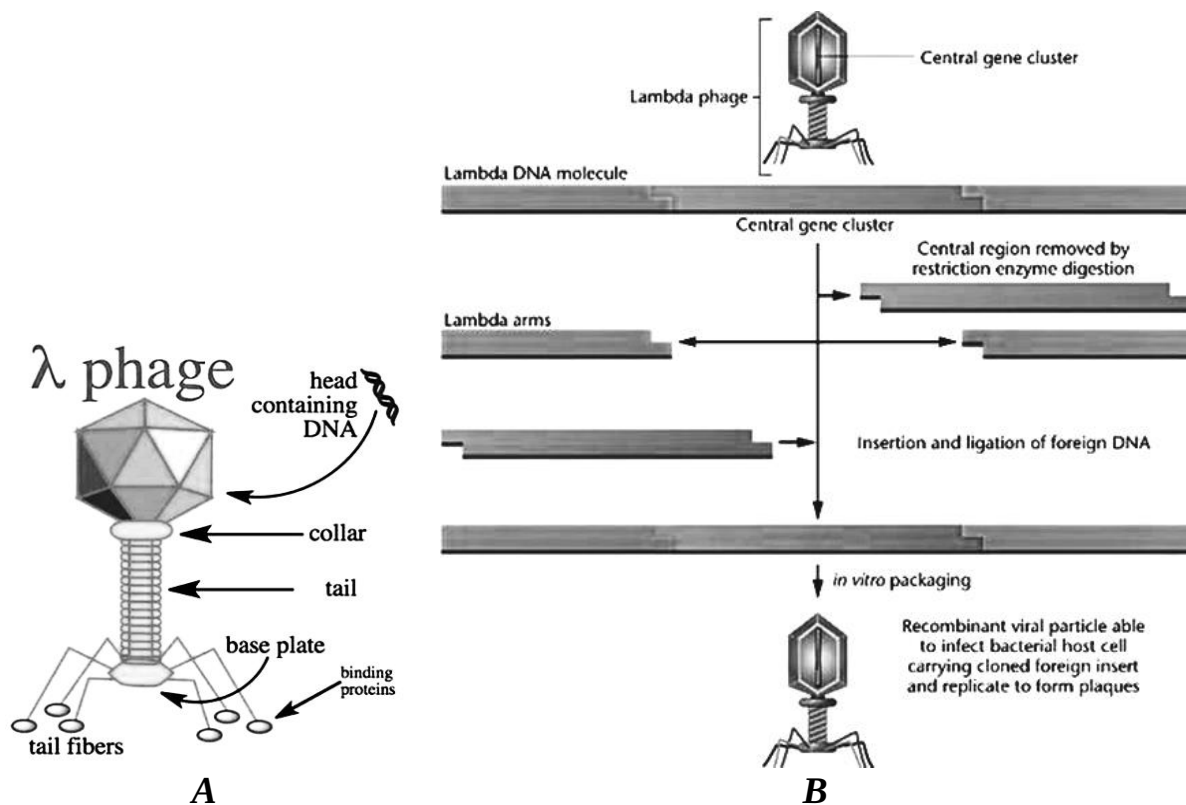


Figure 88. Phage λ :
A — structure; B — recombinant DNA formation based on phage

Vectors based on phage M13. Phage M13 DNA is circular and single-stranded (Fig. 89). This phage recognizes a suitable bacterium by its sex pili, penetrates it, and a copy is synthesized from its DNA matrix: a double-

stranded, replicable form of the phage. M13 does not cause cell lysis but only slows cell development. M13 vectors contain typical selection systems — for example, *lacZα*.

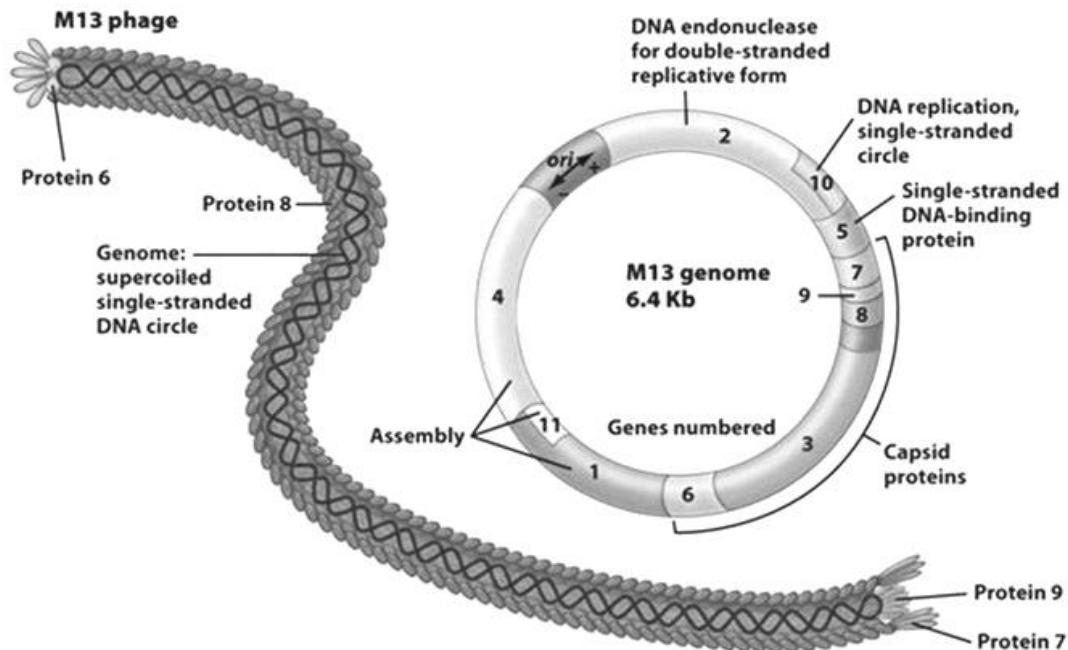


Figure 89. Phage M13 and M13-vector

Single-stranded recombinant DNA is isolated from phages proliferating in *E. coli* cells placed in a nutrient medium, while double-stranded plasmids are isolated as conventional plasmids from the cells themselves.

OTHER VECTORS

Cosmids are artificial structures based on plasmids and phage λ . A cosmid has an ori that allows it to replicate in *E. coli*, an antibiotic resistance gene, a polylinker, and cos sites embedded from phage λ (Fig. 90).

Due to the presence of cos sites, cosmids can be packed into the phage virion. The phage particles allow the recombinant DNA to enter the cell, where it is closed in a ring and replicated as a plasmid. The capacity of cosmids is 35–45 kb.

Phasmids (phagemids) are artificial vectors based on phage λ (or phage f1) DNA and plasmids. They can be used, depending on the conditions, as phages and as plasmids.

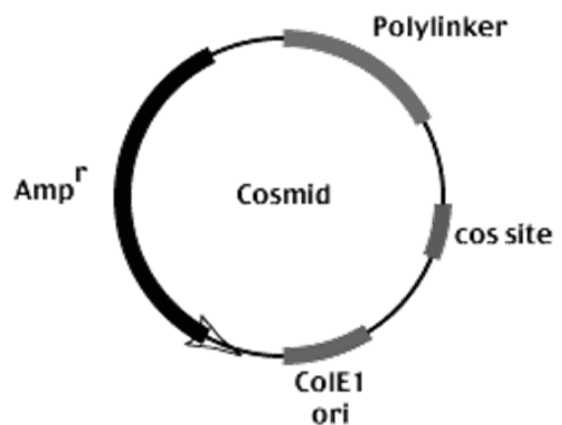


Figure 90. Cosmid

A typical phasmid is 30 kb in size. It contains up to 70 % of the phage genome, including the cos-sites and ori sequence of phage λ , and 30 % of plasmid DNA with the plasmid origin of replication and marker. Phasmid contains all the genes necessary for lytic development of phage λ but is not packed into capsids. The incorporation of 20 kb of foreign DNA in vitro increases the size of the vector, which gives the phasmid the properties of an infectious phage that infects bacterial cells and develops there as a plasmid. The capacity of the phasmid is 20 kb.

Shuttle (binary) vectors are vectors capable of replicating in host cells of different biological species. The need to use shuttle vectors in genetic engineering is because it is more convenient to clone vector DNA in bacterial cells, and obtain biologically active products of cloned genes of higher organisms. An example is the shuttle vector, which can replicate in both *B. subtilis* and *E. coli* cells. Foreign genes are initially cloned in a simpler *E. coli* system, for which genetic engineering techniques have been developed in detail. The selected and modified hybrid plasmids containing two replicons are then transferred into *B. subtilis* cells and the desired protein is obtained. There are shuttle vectors capable of replicating in both bacterial and eukaryotic cells.

Artificial chromosomes are DNA molecules or fragments assembled in vitro from components that guarantee the stable maintenance of large DNA fragments with the properties of natural chromosomes. The capacity of such vectors is 50–300 kb. Types of artificial chromosomes:

- Bacterial artificial chromosome (BAC).
- Yeast artificial chromosome (YAC).
- Mammalian artificial chromosomes (MAC).

RECOMBINANT DNA CONSTRUCTION

One of the methods of constructing recombinant DNA is the restriction-ligation method (Fig. 91):

1. Restriction — cutting the plasmid and the DNA under study by restriction enzyme into fragments with identical “sticky” ends.
2. Ligation — inclusion of DNA fragments into plasmids due to “cross-linking of sticky ends” by enzyme ligase.

Recombinant plasmids are then delivered into bacterial cells and the transformed cells are selected on selective media.

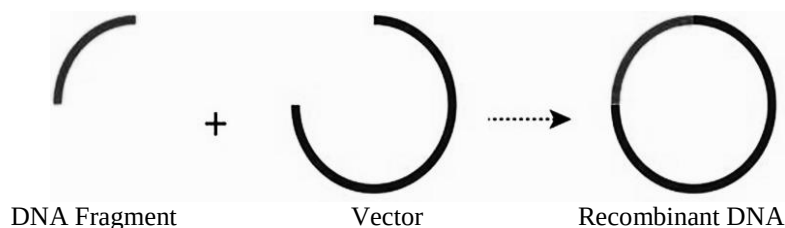


Figure 91. Scheme of the restriction-ligase method for obtaining recombinant DNA

TRANSFERRING RECOMBINANT DNA INTO A HOST CELL

Transformation — the ability of a bacterial cell to take up DNA molecules from a solution. The word “transformation” is also used in a more general sense to describe the uptake of any DNA by any cell.

The yield of transformed cells is 10^{-5} to 10^{-7} of the total number of cells and depends on the strain, the size of the DNA molecules, and the experimental conditions. The efficiency of transformation (up to 10^{-3}) can be increased by various methods.

Transduction is the transfer of recombinant DNA into a bacterial cell by a bacteriophage. During infection, the bacteriophage is adsorbed onto the cell and penetrates it at a rate close to 100 %. This same method is sometimes referred to as transfection. However, for eukaryotic cells (usually animal cells), the word “transfection” also means the introduction of recombinant DNA into the cell by any non-viral method.

Conjugation is the transfer of DNA from a donor cell to a recipient cell. Most plasmids used in work with recombinant DNA are not capable of transferring into cells by conjugation on their own (nonconjugative plasmids). Therefore, they are introduced into cells together with conjugative plasmids. The transfer of nonconjugative plasmids by conjugative plasmids is called mobilization.

Electroporation — a mixture of cells and DNA is subjected to an electric discharge of 2–4 thousand volts, channels are formed in the membrane and DNA penetrates the cell through them (Fig. 92).

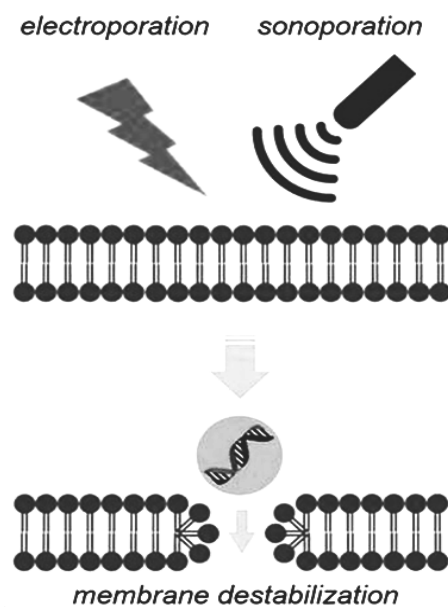


Figure 92. Electroporation

Biolistics (biological ballistics) — DNA bound to inert microparticles (gold, tungsten) is “shot” into target cells using a **gene gun**.

Lipofection — the fusion of the cell membrane with a liposome containing DNA (Fig. 93). Liposomes are spherical vesicles with one or more lipid bilayers. They are formed in mixtures of phospholipids with water. The size of liposomes ranges from 20 nm to 10–50 μm .

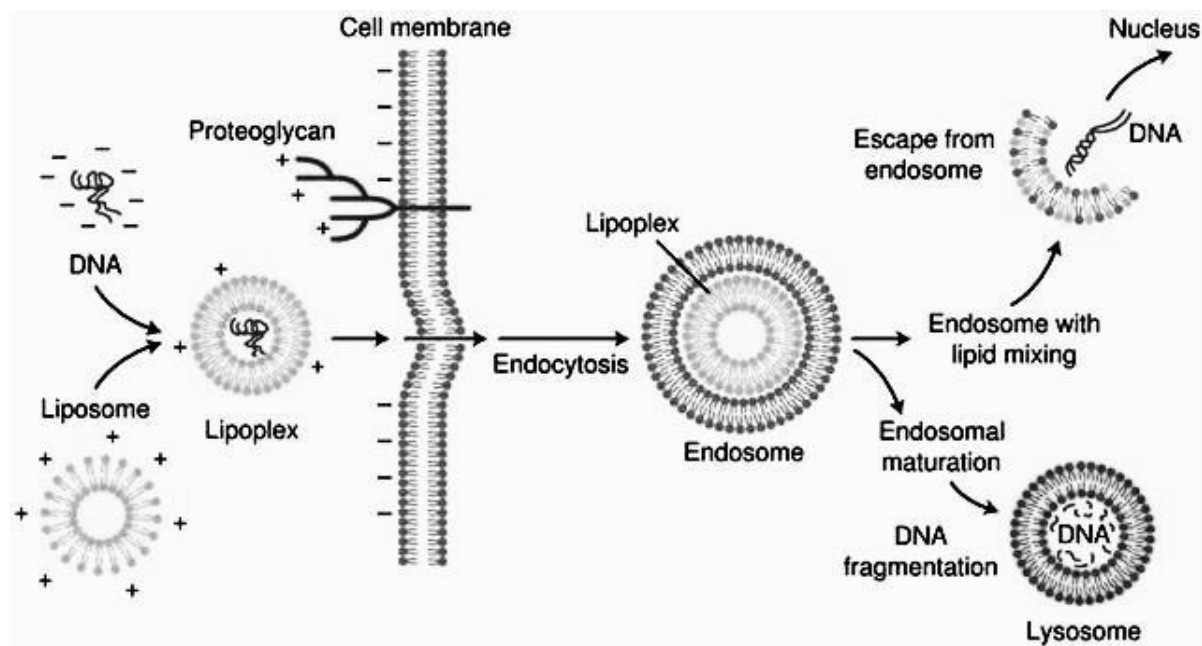


Figure 93. Liposome containing DNA

Microinjection is the direct introduction of DNA into the nucleus of a eukaryotic cell using a micropipette or microinjection needle (Fig. 94).

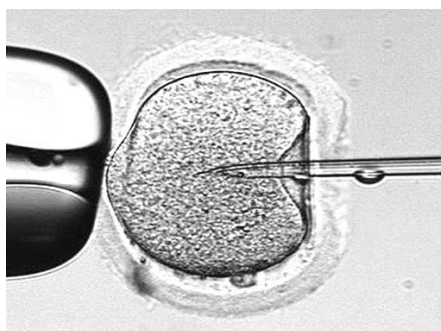


Figure 94. Microinjection of DNA into the cell nucleus. Needle diameter 0.1 μm

BIOTECHNOLOGY, ITS IMPORTANCE FOR MEDICINE

Biotechnology is a discipline that studies the possibility of using living organisms to solve technological problems, as well as the possibility of creating living organisms with the necessary properties by methods of genetic engineering.

Products of biotechnology and genetic engineering in medicine are **recombinant proteins** such as insulin, **antibodies**, **components of vaccines**, and **DNA probes** (see Chapter 9). **Gene therapy** is also a very promising application of genetic engineering technologies in medicine.

GENETICALLY MODIFIED FOODS

Genetically modified foods (GM foods), also known as genetically engineered foods (GE foods) are foods derived from **genetically modified organisms (GMOs)** — plants and animals.

There is a very common misconception that GM foods are different from “natural” food and may even harm humans. But all multiple researches show no difference between GM goods and common foods. Thus, statements about the harm of GM foods have no basis.

Depending on a country’s legislation, foods derived from GMOs, including microorganisms, may also be considered genetically modified.

Plants or animals whose genotype is altered by the introduction of foreign DNA are called **transgenic**, inserted DNA is called a **transgene**, and the process is called **transgenesis**.

GMO PLANTS

Genetic engineering of plants is the transfer of foreign genes into plant organisms, which cause new useful properties of recombinant plants. The creation of transgenic plants for the needs of agriculture solves some practical problems.

1. Obtaining virus-resistant plants. There are various approaches to creating plants that are resistant to viruses. One of them is RNA interference, described in the previous chapter.

2. Obtaining insect-resistant plants. To create insect-resistant plants, the gene for the bacterial Bt toxin that causes insect disease is inserted into the plant genome. In agriculture, Bt-cotton, Bt-rapes, and Bt-corn are used. The main advantage of using Bt plants is an increase in yield and a reduction in the use of herbicides and insecticides, which reduces environmental pollution.

3. Improving the quality of food and agricultural products. Many plant foods contain insufficient amounts of essential amino acids and vitamins. To increase the number of certain amino acids, genes responsible for the biosynthesis of spare proteins are inserted into the plant genome. During their synthesis, the codons of essential amino acids, first of all — lysine, are read more frequently. Varieties of oilseed crops with an altered composition of fatty acids in the fruit (soybean, rape, sunflower) have been created. Currently, agriculture is growing transgenic rice with increased content of carotenoids (golden rice) and transgenic soybeans with an improved protein composition.

4. Improvement of marketable qualities of plants. Changes in taste, appearance, and preservation of fruits are important for the marketing of agricultural products. Various hydrolytic enzymes are activated during fruit ripening. This can make fruits softer and reduce their shelf life. There are transgenic plants in which the synthesis of these enzymes is suppressed, which

improves their marketable quality. This approach has been used to obtain transgenic tomatoes, whose fruits are long-lasting and the plants themselves are resistant to fungal diseases.

5. Obtaining herbicide-resistant plants. One of the technologies to make weed control cheaper is the production of herbicide-resistant cultivated plants. This can be done by introducing a herbicide resistance gene into the plant genome, modifying a herbicide-sensitive protein gene, or modifying herbicide uptake by plants.

6. Obtaining frost-resistant plant varieties. The main damaging agent during freezing is crystalline ice. To prevent its formation, some fish and insects produce special hydrophilic proteins. The genes of these proteins are transferred to plants, and their frost resistance increases significantly.

7. Plant bioreactors are plants capable of synthesizing: insulin, antibodies, and other proteins for medical needs. This helps solve the problem of proper modification of eukaryotic proteins, which is not possible in bacterial cells.

TRANSGENIC ANIMAL BIOTECHNOLOGY

Transgenic animal biotechnology is a rapidly developing area of genetic engineering. Its methodology was developed on laboratory mice. The introduction of foreign DNA into an animal is performed using retroviral vectors, microinjection of DNA, stem cells, nucleus transfer, and artificial chromosomes.

The use of transgenic animals. Genetically modified laboratory animals have for many years remained an effective tool for studying the functional properties of genes, proteins, and other molecules. Using such animals, it is possible to study the pathogenesis and molecular features of diseases, identify and validate new therapeutic targets, and conduct effective search and development of new drugs, including preclinical studies. At the same time, genetically modified animals are also becoming increasingly attractive in fields such as animal breeding, where genome modifications can be used to correct economically valuable animal traits.

Finally, transgenic animals can serve as bioreactors for the synthesis of recombinant proteins secreted into milk, allowing the production of recombinant proteins in significantly larger quantities and at a much lower cost compared to the production of proteins in eukaryotic cell cultures.

Examples of transgenic animals and the proteins they produce:

- Cattle: growth hormone, lysostaphin, albumin.
- Goats: alpha-fetoprotein, granulocyte colony-stimulating factor, growth hormone, antithrombin,
- Mice: clotting factor IX, lysostaphin, CD6-receptor, S-glycoprotein of gastroenteritis virus sheath, hepatitis A virus sheath protein antigen.
- Rabbits: erythropoietin, blood clotting factor VIII.

CHAPTER 9 GENETICALLY ENGINEERED PRODUCTS

METHODS OF MONOGENIC DISEASE DIAGNOSIS

Various molecular genetic tests are used to diagnose genetic diseases in a fetus, children, and adults and to identify carriers, as well as to detect genetic diseases in embryos created by in vitro fertilization. For adult genetic testing, DNA from white blood cells is usually used. Alternatively, many genetic tests can be performed on cheek epithelial cells from oral smears or on hair cells. Some genetic tests can be performed on gametes.

PCR-based methods. Over the past decades, many PCR-based methods for detecting already known mutations have been developed. The principle of PCR is described in Chapter 7.

PCR product size analysis. In some cases, a mutation can be detected simply by estimating the length of a PCR-amplified DNA fragment by electrophoresis in agarose gel. The length can change as a result of deletions and insertions in the gene. For example, the vast majority of Friedreich's ataxia cases are caused by multiple increases in the number of copies of the GAA repeat in the first intron of the FXN gene. As a result, the length of the DNA fragment containing the repeat increases, and mobility in the agarose gel decreases.

RFLP. A classical method of genetic testing was restriction fragment length polymorphism (RFLP) analysis. A sufficiently long DNA molecule can randomly contain restriction sites for some endonucleases. When identical DNA sequences are cleaved with a particular restriction endonuclease, the same set of DNA fragments containing fragments of different lengths will be obtained each time. If they are separated by gel electrophoresis, a similar set of fractions of a certain length will be observed each time. However, if one pair of nucleotides has changed in the analyzed DNA sample, and it happened just at the restriction site, the DNA fragments obtained by the corresponding restrictase will change (Fig. 95).

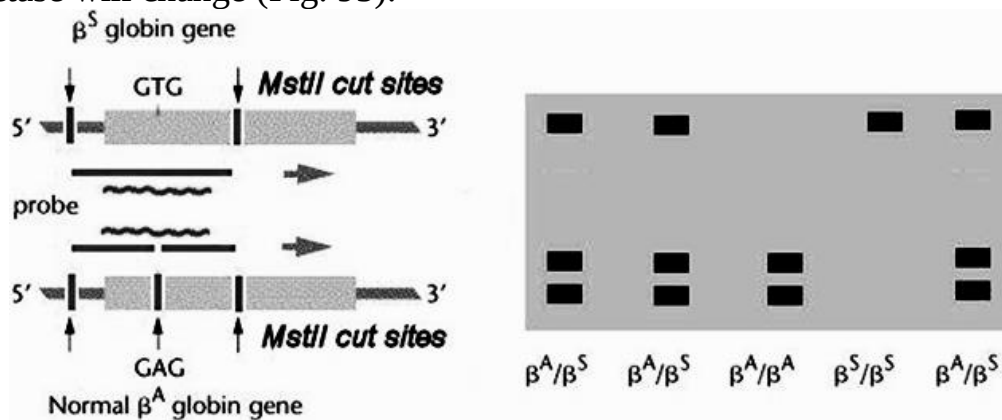


Figure 95. Schematic application of restriction fragment length polymorphism for the diagnosis of sickle cell anemia

SSCP. The SSCP (Single-strand conformational polymorphism) assay is a relatively simple and sensitive technique that also allows the detection of mutations through gel electrophoresis. The principle of SSCP analysis is based on the fact that single-stranded DNA fragments acquire a certain three-dimensional conformation due to interactions between different sections of the same molecule. This conformation depends on the nucleotide sequence of the DNA strand and differs even for fragments of the same length. In addition, it affects the electrophoretic mobility of the fragment. A change in conformation caused by a change in one nitrogenous base can lead to a change in the migration rate of single-stranded DNA in gel electrophoresis, so samples of nonmutant and mutant DNA will show different band arrangements. To perform SSCP, the required DNA sequence is first amplified using PCR, double-stranded DNA is denatured, cooled to maximize nucleotide-binding of the same strand, and then electrophoresis is conducted.

Allele-specific PCR. Allele-specific PCR is two reactions in different tubes. One uses a primer complementary to the normal sequence and the other — complementary to the mutant sequence. The second primer added to both tubes is the same (recall that PCR requires two primers marking the boundaries of the amplified fragment). In the presence or absence of a mutation in the studied DNA, only the fragments in the tube with the corresponding primer will be effectively amplified.

Real-time PCR. The principle of real-time PCR is described in Chapter 7.

USING DNA MICROARRAYS

The working principle of DNA microarrays was previously discussed in Chapter 7. Some genetic diseases are caused by changes in a single gene, while others are characterized by changes in many of them. Many different single-nucleotide polymorphisms can be represented on a microarray. In this way, it is possible to test the DNA under study simultaneously for thousands of mutations that can both cause individual diseases and create a predisposition to cardiovascular disease, asthma, diabetes, Alzheimer's disease, and other genetically determined pathologies.

Sequencing. Today, the human genome can be sequenced in just several days at a cost of about \$1000. Therefore, it is possible to sequence individual cDNAs and exons for diagnostic purposes, looking for mutations in certain genes, or the entire genome if necessary. Sequencing of exomes, i.e. sequences corresponding to exons of protein-coding genes, can serve as an alternative to whole-genome sequencing since most of the mutations causing genetic diseases occur precisely in them. In the near future, whole-genome sequencing, exome sequencing, and RNA sequencing should be expected to replace DNA microarrays and make this technology obsolete.

PRODUCTS OF GENETIC ENGINEERING IN MEDICINE

A genetically engineered approach underlies the production of many products used in medicine and biotechnology.

Proteins used in the therapy of certain diseases. The classic example of recombinant proteins is insulin — pancreatic hormone regulating carbohydrate metabolism. Previously, it was isolated from the pancreas of pigs and cattle. The disadvantages of this method were some differences in the amino acid composition of animal insulin from human insulin and the presence of some other impurity proteins, which made such insulin products allergenic. In addition, about 1000 kg of the pancreas could be required to produce 100 g of insulin (one cow gland weighs about 200–250 g). Currently, recombinant human insulin is mainly produced by *E. coli* or baker's yeast cells. Other proteins are also produced in microorganisms or animal cell cultures: Somatotropin (growth hormone produced by the anterior pituitary lobe), erythropoietin (kidney hormone that stimulates erythropoiesis — the formation of new red blood cells), interleukins (proteins synthesized mainly by white blood cells as immune system signaling molecules), anticoagulants (prevent the formation of blood clots, used, for example, in myocardial infarction), clotting factors VIII and IX (used for replacement therapy of hemophilia), interferons (proteins secreted by cells in response to virus invasion and stimulating the antiviral defense of neighboring cells), etc.

Vaccines. A vaccine is a medication that provides the body with acquired immunity to a certain disease. This is achieved by presenting the antigens of the pathogen of a certain disease to the immune system. There are different types of vaccines, among them:

- Attenuated vaccine or a live attenuated vaccine, LAV (contains a weakened pathogen).
- Inactivated vaccines (contain a “killed” pathogen).
- Subunit vaccines (contain individual proteins of the pathogen or their fragments, or other components).
- Vector vaccines (viral vectors whose genome contains sequences encoding fragments of the pathogen).
- mRNA vaccines (contain individual mRNAs encoding pathogen fragments).

The components of subunit vaccines can be obtained from the pathogens themselves or by recombinant DNA technology. In the second case, the vaccines would be called recombinant vaccines. An example would be the hepatitis B vaccine containing the surface antigen of the virus produced in genetically engineered yeast cells. In contrast to recombinant vaccines, vector, and mRNA vaccines provide the production of the pathogen antigens by the human cells themselves, located primarily at the site of vaccine

administration. mRNA vaccines deliver the mRNA encoding the pathogen antigen to the cell. Known examples are some COVID-19 vaccines containing mRNA encapsulated in lipid nanoparticles.

Vector vaccines are viral vectors containing sequences in their modified genome that ensure the synthesis of certain antigens by the cell. Certain COVID-19 vaccines are also known examples. They are based on adenoviral vectors that do not integrate into the genome and cannot reproduce. As with mRNA vaccines, cells predominantly located at the site of injection, into which the necessary gene has been delivered, start producing the viral protein that triggers the immune response.

Antibodies. Antibodies (immunoglobulins) are soluble proteins that can bind to an antigen (e.g. a protein on the surface of a virus or bacteria). The main type of antibodies in the blood and intercellular fluid are immunoglobulins G (IgG). Antibodies are produced by plasma cells (mature B-lymphocyte) as a result of an encounter with a corresponding antigen (e.g. in case of disease), with each cell producing only one type of antibody to the same antigen site. The main function of immunoglobulins is to “recognize” the antigen and bind to it, which, firstly, “marks” the antigen carrier as to be utilized and, secondly, can bind parts of the antigen with specific functions, preventing their fulfillment. For example, antibodies that bind to a protein necessary for the entry of the virus into the cell reduce the ability of the virus to penetrate the cell. Antibodies have a Y-shaped structure and consist of four polypeptide chains (two heavy and two light) linked by disulfide bonds (Fig. 96).

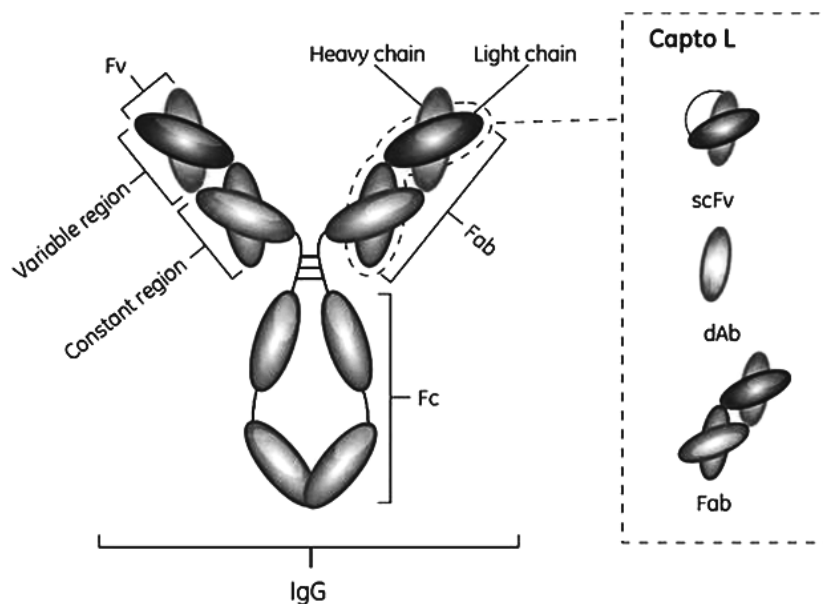


Figure 96. Scheme of the antibody structure

Two of the three “ends” of the immunoglobulin contain the variable part of the antibody, which is responsible for the recognition and binding to the corresponding antigen, while the rest of the antibody is constant, being

the same for all antibodies of this type. Variable regions are present on both heavy and light chains. Within the variant regions, there are hypervariable regions, regions that are in direct contact with the antigen.

Antibodies are of great interest in the diagnosis and treatment of diseases because of their ability to bind selectively to a specific antigen. Antibodies used for medical purposes are divided into monoclonal and polyclonal antibodies.

Polyclonal antibodies (pAbs) are a heterogeneous mixture of different IgG against different regions of the antigen and are produced by different plasma cells. They are obtained by injecting a specific antigen into laboratory animals such as rabbits, goats, etc.

Monoclonal antibodies (mAbs) are molecules of the same IgG against the same region of the antigen and are produced by the same cell (or its progeny).

One of the laboratory methods for producing monoclonal antibodies is hybridoma technology (Fig. 97).

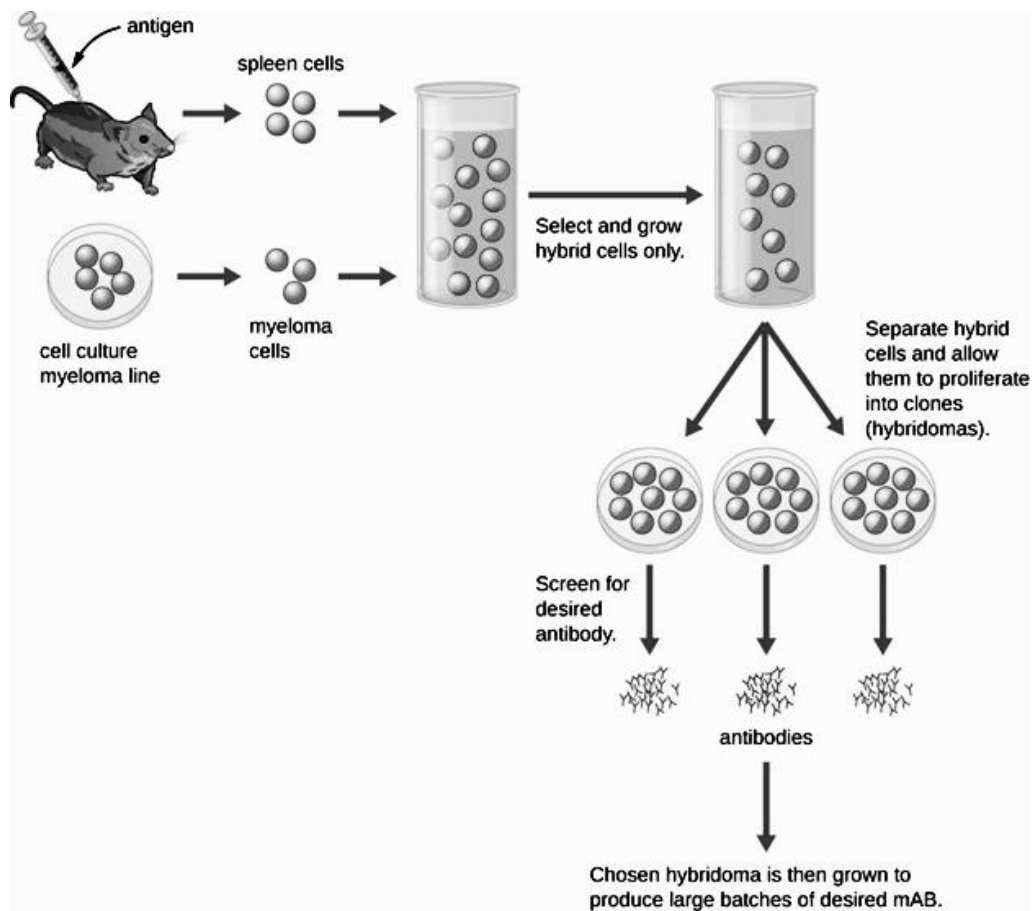


Figure 97. Hybridoma technology

This method consists in fusing B-cells producing the necessary type of antibodies, obtained from the spleen of immunized animals, most often mice, with myeloma cells (cancer developing from plasma cells). Malignant cells, unlike normal cells, have no limit to the number of divisions. The fusion of the

two cell types results in hybridomas — cells that produce certain antibodies and can divide an unlimited number of times like tumor cells. Once the desired hybridomas have been selected, they are cultured on nutrient media to produce cell lines that produce the necessary monoclonal antibodies of animal origin (murine monoclonal antibodies).

However, murine monoclonal antibodies are often ineffective in therapy because of the high immunogenicity of foreign murine proteins. Genetic engineering was used to overcome the problems. Instead of murine antibodies, other types of antibodies have been obtained (Fig. 98):

- Chimeric monoclonal antibodies.
- Humanized monoclonal antibodies.
- Fully Human monoclonal antibodies.

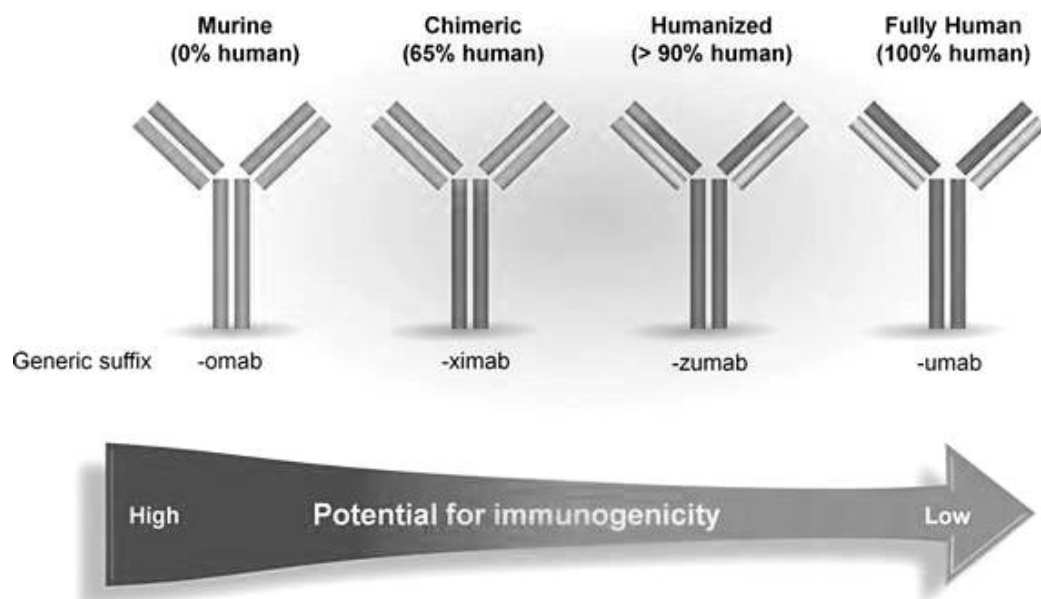


Figure 98. Types of antibodies

IN VIVO GENOME EDITING SYSTEMS

CRISPR/Cas9. CRISPR/Cas9 is a protein and RNA complex that allows DNA editing in vivo (i.e., inside a living cell). The protein of this complex, Cas9, is a nuclease capable of cutting both strands of DNA. The cutting region is found by a Cas9-associated RNA molecule whose site “recognizes” complementary nucleotide sequence. This molecule is called single-guide RNA (sgRNA). There is also another condition without which DNA will not be cut: the presence of a special three-nucleotide sequence called PAM (protospacer adjacent motif) next to the region that is bound to sgRNA (Fig. 99).

In nature, such a complex ensures inherited “antiviral immunity” in bacteria, which helps to recognize foreign phage DNA using complementary RNA and to cut it up using Cas proteins. The matrix for the synthesis of such

RNA is the fragments of phage DNA stored in the bacterial genome, which it or its ancestor had encountered previously. These fragments are embedded in a special region of the genome called CRISPR (clustered regularly interspaced short palindromic repeats), which allows the bacterium to “remember” the phage and respond to its invasion in the future.

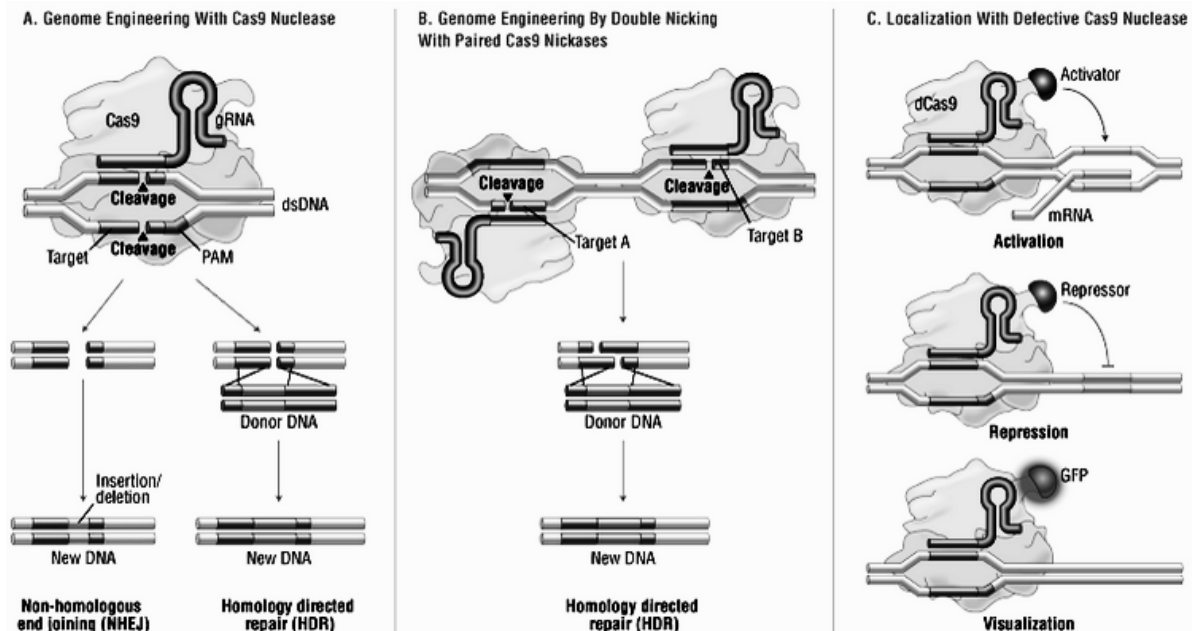


Figure 99. Applications of CRISPR/Cas9

In genetic engineering, the necessary sgRNA guide sequence (about 20 nucleotides) is prepared to affect any DNA region of interest, provided it is adjacent to the PAM. Then, the DNA encoding Cas9 and sgRNA is introduced into the plasmid and delivered to the cells to be edited.

Similar to natural CRISPR/Cas9 introduces a double-stranded break in the recognized DNA strand. More often than not, such breaks are repaired in the cell by an imperfect method called nonhomologous end joining (or NHEJ). This mechanism often causes nucleotide insertions or losses when a gap in the DNA is “glued” together. This, in turn, often leads to a frameshift (i.e., a change in the codon start point at the site following such mutation), causing a loss of gene function. However, with some probability, a repair can occur by another mechanism: an intact copy of the edited DNA region (for example, the same gene in a homologous chromosome) can be used as a template. This mechanism is called homology-directed repair (HDR). Therefore, if it is supposed to replace the edited DNA region with some other, the necessary sample is also delivered to the cell: single- or double-stranded DNA containing, first, regions complementary to the regions located on either side of the editing site, and, second, the desired sequence between them.

Various modifications of CRISPR/Cas9 allow other tasks to be performed:

- Nitrogenous base editing. Cas9 is deprived of nuclease activity and binds to an enzyme that converts cytosine to uracil, perceived by the cell as thymine, or adenine to hypoxanthine, perceived by the cell as guanine. Thus, in certain cases, it becomes possible to change the base pairs TG to TA or AT to GC.

- Introduction of single-strand breaks. The modified Cas variant cuts a single strand of DNA. If two complexes that recognize adjacent sequences on two different DNA strands separately are used, the chance of non-targeted cutting of the DNA can be reduced. This is since the break-in of two strands at once will be obtained only if both complexes recognize the target regions.

- Transcription activation or repression. Cas9 devoid of nuclease activity is coupled to a transcription activator protein and directed to the promoter or regulatory regions of a particular gene. Similarly, a repressor protein can be attached to Cas9 to inhibit the transcription of a particular gene.

- Epigenetic modifications. Cas9 deprived of nuclease activity is coupled to proteins that carry out epigenetic modifications. Available modifications include histone acetylation and demethylation, and cytosine methylation and demethylation.

- Visualization. Catalytically inactive Cas9 is combined with a fluorescent marker such as green fluorescent protein (GFP) to visualize target loci in the genome. This method is easy to implement because of the simplicity of sgRNA construction and makes it possible to detect chromatin dynamics in living cells.

Despite its broad capabilities, the use of CRISPR/Cas9 has some limitations, including:

- In theory, the CRISPR/Cas9 system should have extremely high accuracy, but in reality, it often exhibits non-targeted activity, recognizing a large number of partially complementary sgRNA sites. For this reason, improvement of this system is ongoing, and its application is still limited.

- When a region of DNA is edited, it is not always repaired by homologous recombination, allowing the necessary DNA sequence to be “inserted” into the genome.

- Another significant limitation is the large size of CRISPR-Cas9, which makes it difficult to deliver to cells in vivo.

The CRISPR/Cas9 system is currently being improved, and some works demonstrating improvements in its accuracy and efficiency suggest that the existing limitations are not insurmountable.

CHIMERIC NUCLEASES

In addition to CRISPR/Cas9, there are other genome editing systems: zinc finger nucleases (ZFN), and transcription activator-like effector nucleases

(TALEN). Unlike Cas proteins, they do not bind to guide RNAs and recognize a specific site in DNA due to the structure of their DNA-binding domains.

Before CRISPR/Cas, chimeric nucleases were developed for genome editing, i.e., proteins “assembled” from two structural units, one of which should recognize exactly the DNA site to be cut and bind to it, while the other should cut that site.

Like CRISPR/Cas, chimeric nucleases can be produced directly in the cell by introducing vectors containing the corresponding genes. Such vectors are provided with a nuclear localization signal to ensure that the genome editor enters the nucleus.

Zinc-finger nucleases (ZFNs) were the first among chimeric nucleases. They comprise protein domains containing a zinc atom, called zinc-finger nucleases. Each of these domains is capable of specifically binding to a specific three-nucleotide sequence. However, this genome editing tool is quite often bound to non-target sites, as well as being difficult and expensive to develop and, therefore, not widely used.

TALEN (Transcription Activator-Like Effector Nucleases) technology turned out to be more promising. Such nucleases have DNA-recognizing protein domains, each of which recognizes only one nucleotide. Proteins (TAL-effectors) of special bacteria parasitizing in plant cells served as prototypes of such domains. The role of these proteins is to act as transcription factors and bind to specific DNA sites, activating the genes necessary for the survival of the parasite.

Since in this case one nucleotide is recognized by a single protein domain, obtaining a construct that recognizes the desired nucleotide sequence is relatively simple. It is combined with a DNA-cutting enzyme (usually FokI). FokI functions only as a dimer (a complex of two subunits of this enzyme), so two constructs containing it are required for DNA cutting — each of them recognizes its specific site on different strands of the DNA molecule. This independent recognition of sites on the two strands reduces the chance of untargeted recognition and increases the accuracy of the system.

The emergence and active use of genome editing systems open up some new possibilities for humans. These technologies can be used both on humans (to treat tumors, genetic, and other diseases) and other living organisms (improving food quality, and fighting pathogens and disease vectors). The promise of such methods is well demonstrated by the example of gene therapy. However, this also involves some ethical issues. For example, genome editing implies use for therapeutic purposes, but theoretically, these techniques could also be used for other purposes, such as genome improvement. As public health focuses on prevention, the line between disease prevention and human enhancement is difficult to define.

Therefore, such issues need to be well-defined and regulated by a sufficiently robust and dynamic regulatory framework.

GENE THERAPY

Gene therapy is an approach to disease therapy aimed at manipulating gene expression or changing the biological properties of living cells. For this purpose, a normal copy of a mutant gene can be injected into a patient, or a gene associated with the development of pathology can be inactivated.

The basis of the technology is the introduction of the necessary genetic material into a vector and the further transfer of the vector into the cell.

Theoretically, gene therapy can be performed both on somatic cells and on germline cells (gametes and their precursors). However, genome modification of germline cells is not currently performed and is generally forbidden by law due to the ethical issues involved and the lack of accuracy of the available genome editing technologies. Therefore, gene therapy is performed on patients' somatic cells, and, hence, the changes made in the genome of the target cells are not passed on to subsequent generations.

Gene therapy is also performed *in vivo* and *ex vivo*. *In vivo* gene therapy involves the introduction of genetic material directly into the human body, where the vector reaches the target cells and penetrates them, thereby changing their properties. In the case of *ex vivo* gene therapy, the target cells are first obtained from the patient's body, then the necessary genetic material is injected into them in the laboratory (Fig. 100).

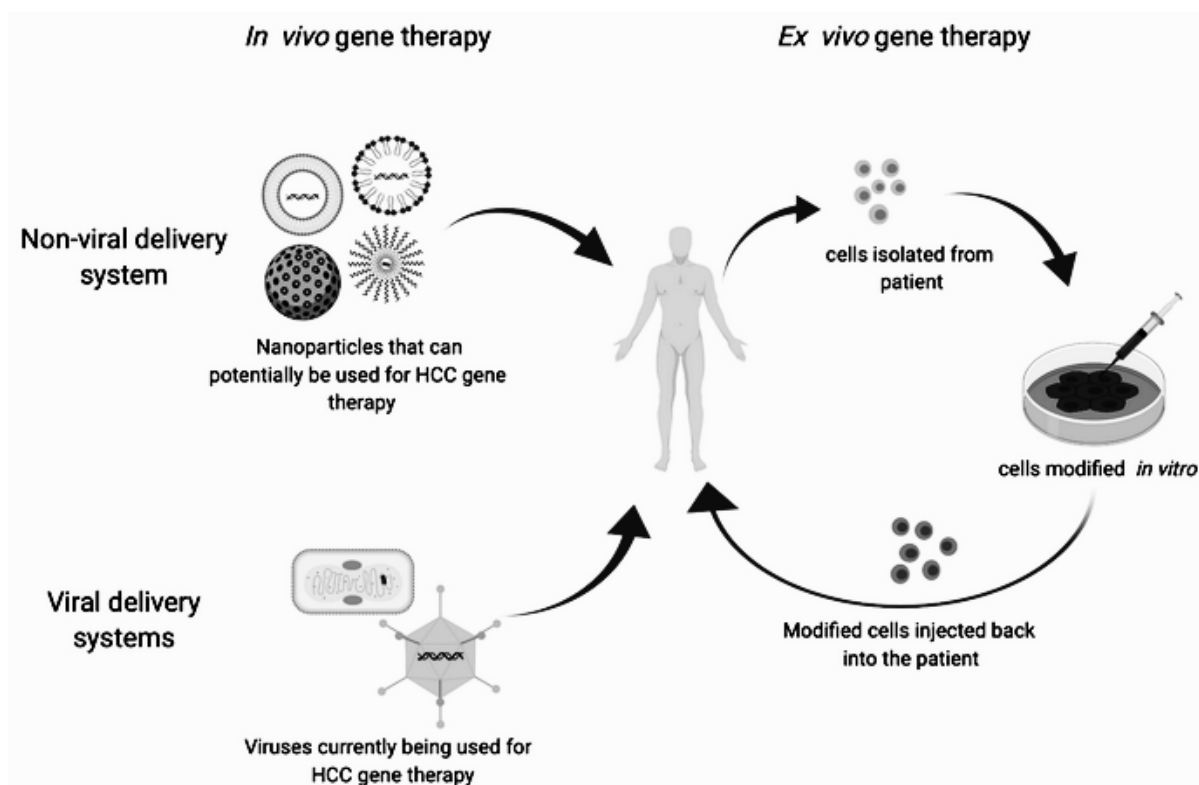


Figure 100. Principles of *in vivo* and *ex vivo* gene therapy

The successfully altered cells are selected, allowed to multiply on a nutrient medium, and then injected back into the patient.

As a rule, viruses are used as vectors because they have the natural ability to deliver their genetic material into cells. Retroviruses, lentiviruses, adenoviruses, and adeno-associated viruses (AAV) are usually used. The choice of such a vector for a particular situation is related to their peculiarities since each of them has both advantages and disadvantages. For example, the advantages of AAVs are that they are non-pathogenic, can be relatively easily targeted to different cell types, and can infect both dividing and non-dividing cells. At the same time, they are unable to carry large fragments of DNA and usually do not integrate into the genome, which prevents them from being passed on to all progeny of dividing cells. Integration into the genome is typical of retroviral and lentiviral vectors, but, at the same time, it is associated with some risk of insertional mutagenesis, which can lead to activation of protooncogenes and rearrangement in tumor cells.

As of 2021, gene therapy products against the following diseases have been approved or are undergoing clinical trials:

- Severe combined immunodeficiency (SCID) caused by adenosine deaminase deficiency (ADA deficiency) was the first disease against which gene therapy using γ -retroviral vectors was first used in 1990. There is also an X-linked variant of this disease, which is caused by mutations in one of the cytokine receptor subunits. Its gene therapy with γ -retroviral vectors was also initiated in the late 1990s. Most children who received therapy showed signs of complete immune system recovery. However, a few patients eventually developed T-cell leukemia. In 2016, a new drug, Strimvelis, for ex vivo therapy of ADA-SCID was approved for use by the European Medical Agency (EMA).

- β -hemoglobinopathies — β -thalassemia and sickle cell anemia. These diseases are caused by mutations in the HBB gene encoding the beta chain of hemoglobin. In 2019, a medication based on the lentiviral vector was approved in Europe.

- A form of retinal dystrophy (caused by mutations in the RPE65 gene). An AAV vector-based medication for in vivo gene therapy of this pathology (Luxturna) was approved in the United States in 2017.

- Spinal muscular atrophy is a neurodegenerative disease caused by mutations in the SMN1 gene. The protein encoded by this gene is necessary for the survival of motor neurons. In 2019, Zolgensma, a product carrying the SMN1 gene in an AAV vector, was approved in the United States.

- Cerebral adrenoleukodystrophy (CALD) — X-linked disease associated with peroxisome dysfunction and affecting the nervous system. In 2023, this medicine is the most expensive one in the world (costing about \$3 million).

- Lipoprotein lipase deficiency is a disease caused by mutations in the LPL gene. In 2012, the product Glybera, which contains this gene as part of an adeno-associated viral vector, was approved in the countries of Western Europe.

- Hemophilia B is a disease caused by mutations in the clotting factor IX gene. In 2018, clinical trials were initiated for a product based on an AAV vector carrying the “Padua” variant of this gene under the control of a liver cell-specific promoter. The clotting factor encoded by this variant has several times more activity than the “normal” protein.

- Tumor therapy includes a variety of approaches. Among the first were injections of an AAV vector, with a gene that increases the sensitivity of tumor cells to certain medications, or with a tumor suppressor gene encoding the p53 protein, one effect of which is to trigger apoptosis.

Another approach is to use oncolytic viruses, which multiply in tumor cells but do not affect healthy cells. In the treatment of some oncohematological diseases, chimeric antigen receptor T-cell (CAR-T) therapy is used. It is described later in this chapter.

BIOINFORMATICS

Modern biology deals with extremely large amounts of data that cannot be processed without computer methods. For example, a sequenced genome can be many thousands of short fragments (often several hundred nucleotides each) that need to be put together, which is unacceptable to do manually. The human genome contains over three billion base pairs, and analyzing it without computer methods would be extremely difficult.

Computer usage for solving biological problems is called bioinformatics. There are several main directions in bioinformatics:

- Bioinformatics of sequences (the study of DNA, proteins (Fig. 101), and RNA).

- Structural bioinformatics (studying the structure of proteins).

- Expression analysis (analysis of gene expression and regulation).

- Study of cellular organization (microscopy and image analysis, analysis of protein arrangement, study of chromatin organization).

- Systems biology (analysis and modeling of complex interactions in living systems).

Bioinformatics deals with the analysis of amino acid and nucleotide sequences of proteins and nucleic acids (Fig. 101).

Such sequences are usually written in the form of one-letter notations (for DNA and RNA — A, G, C, T/U; for amino acids there are also one-letter notations). Examples of tasks in this section of bioinformatics include searching the genome for sequences encoding proteins or RNA, searching for binding sites with regulatory proteins, etc. Recall that even the size of bacterial

genomes is usually in the millions of base pairs. By detecting a gene and finding out its nucleotide sequence, it is possible to determine the amino acid sequence of the protein it encodes. Knowing this sequence, it is possible to compare it with known ones. If a similarity with a known protein is detected, then we can assume with a certain degree of probability that the protein encoded by the detected gene performs a similar function. To compare two or more sequences, they are aligned — they are placed under each other in a way that makes it easy to identify similar or different regions in them.

Figure 101. Amino acid sequence alignment of the histone H1 site of a human, chimpanzee, mouse, rat, and cow

The number of known sequences is much larger than the number of experimentally determined structures since laboratory procedures for structure determination are much more complex and expensive. Examples of structural bioinformatics tasks are the prediction of protein sites important for a particular protein function, comparative analysis of structures of related proteins, analysis of structures of protein complexes with other molecules (including other proteins), prediction of the effects of molecules of various substances on protein molecules (e.g., search for potential drugs affecting target proteins), prediction of protein structure from the structure of a protein with a similar sequence.

Internet databases containing information about nucleotide sequences, servers, and specialized online services, Blast, NCBI. Nucleotide and amino acid sequences, information about them, and some tools for working with them are available on specialized online databases.

NCBI — The National Center for Biotechnology Information (USA) is part of the U.S. National Library of Medicine (NLM), an affiliate of the National Institutes of Health (NIH). NCBI was founded in 1988. NCBI contains some databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics. Major databases include GenBank containing DNA sequences, PubMed containing a bibliographic database, BLAST bioinformatics tool, and other resources.

UniProt is a publicly available database of protein sequences that also contains a vast amount of information about their biological functions derived from the scientific literature. It is maintained by the UniProt Consortium, which includes several European bioinformatics organizations and a foundation in Washington, D.C., USA.

GenBank is a database of all publicly available nucleotide sequences and their corresponding amino acid sequences. As of August 15, 2024, GenBank contained a total of 4,500,000,000 sequences consisting of 30,000,000,000,000 bases.

Protein Data Bank (PDB) is a database of three-dimensional structural large biological molecules such as proteins and nucleic acids. The data is freely available from biologists and biochemists around the world, using a variety of methods. The PDB refers to an organization called the World Protein Data Bank, wwPDB.

BLAST (basic local alignment search tool) is an algorithm and program for comparing amino acid or nucleotide sequence information. It functions on the NCBI platform. BLAST search allows the researcher to compare a protein or nucleotide sequence with those already in the database and find similar ones. For example, after discovering a previously unknown gene in an animal, a researcher can find out if the human has a similar gene using BLAST.

MOLECULAR PHYLOGENETICS

Nucleotide sequence analysis also plays an important role in phylogenetics.

Phylogenetics is a field of systematics that deals with the evolutionary history of taxonomic groups of organisms. Evolutionary relationships among groups of organisms are represented as phylogenetic trees. A phylogenetic tree is a diagram showing the lines of the evolutionary origin of different species, organisms, or genes from a common ancestor. Phylogenetic trees can be constructed both from morphological traits and molecular biology data. The basic idea underlying molecular phylogenetics is that evolutionarily closer

species, on average, have fewer differences in the structure of their DNA and proteins than unrelated species.

If several biological species have the same feature in the structure of a certain protein, the random occurrence of which is unlikely, it should be explained by the fact that it was present in their common ancestor and not independently arisen as a result of mutations in each of them. Since the sequences obtained from different species have a large number of similarities and differences, special statistical methods are used to evaluate them and construct phylogenetic trees (Fig. 102).



Figure 102. Hypothetical aligned sequences of some protein regions of several biological species (right) and the phylogenetic tree constructed from them (left)

The most widely used object for phylogenetic analysis is the ribosomal RNA (rRNA) sequences of the small ribosomal subunit (sequences of 16S ribosomal DNA for prokaryotes and 18S ribosomal DNA for eukaryotes). This is primarily because this sequence is present in all organisms, which makes it possible to compare even completely morphologically dissimilar organisms, and also because this sequence is rather conservative (i.e., it is little changed in the course of evolution).

Briefly, the phylogenetic analysis can be divided into five stages:

1. Preparation of sequence data.
2. Sequence alignment.
3. Choice of phylogenetic reconstruction method.
4. Determination of the best phylogenetic tree.
5. Estimation of the phylogenetic tree.

PERSONALIZED MEDICINE

It is now well established that the same treatment can have different efficacy in different patients. However, thanks to the increasing availability of laboratory techniques, it is now possible to take into account the individual characteristics of each patient. Personalized medicine (or precision medicine) can be defined as an individualized, molecular approach to the diagnosis and treatment of diseases. It is an approach that takes into account the individual

characteristics of the genome, proteome, gene expression, and other molecular characteristics of the patient and applies this information to select precise methods of disease treatment and develop new therapies and medications.

Pharmacogenomics. Pharmacogenomics is the study of the genome's role in the response to medications.

Each year millions of people worldwide experience serious side effects from medications. On the other hand, many patients respond worse than expected to pharmacological treatment, which is partly due to their genetic characteristics. Given this, one of the main goals of pharmacogenomics is to examine a patient to identify his genetic characteristics, based on which an individual dosage of a particular medicine is set. There are dozens of genes encoding proteins that affect drug metabolism (receptors, transport proteins, enzymes that degrade or modify drugs).

For example, people with certain cytochrome P450 gene variants metabolize and excrete medications relatively slowly, which can lead to accumulation and overdose. Other people may have genetic traits that lead to the rapid excretion of medications, which reduces their effectiveness. A prime example is the CYP2D6 gene, which encodes an enzyme that is involved in the metabolism of about 25 percent of all pharmaceutical products, including paracetamol, beta-blockers, and others. There are more than 70 allele variants of this gene, some of which decrease the activity of the encoded enzyme and others that increase it.

Another goal of pharmacogenomics is to develop drugs that target the genetic profiles of specific subgroups of patients. This is particularly relevant in the treatment of cancer since each tumor is genetically unique. One example of a successful application of this approach is the treatment of breast cancer and some other cancers with monoclonal antibodies against the HER2 protein. This protein is a membrane-bound receptor for epidermal growth factor (a protein that is a signal for cell division). In about 20 % of all breast cancers, the HER2 gene is amplified and the amount of the receptor on the cell surface is increased. Monoclonal antibodies that bind to HER2 inhibit its signals and cause the cell cycle to stop, destroying the cancer cell. Since this treatment affects only cancer cells with amplified HER2, it is important to know the status of this gene in each case. Immunohistochemistry (see Chapter 1, cytology methods) and FISH (fluorescence in situ hybridization, see Human Genetics) can be used for this purpose.

This example is not the only one. About 40 % of colorectal cancer patients respond to preparations based on monoclonal antibodies to epidermal growth factor receptor (EGFR) because in these cases the cancer cells express EGFR on their surface and also have no mutations in the KRAS gene encoding a protein that is involved in the transmission of signals from this receptor to the nucleus.

Personalized Medicine in Oncology. In oncology, personalized medicine could potentially achieve the use of targeted treatments that address specific mutations and gene expression defects in each tumor, leading to remission and cure. To date, the percentage of patients who can be successfully treated in this way is small, but major research programs, such as The Cancer Genome Atlas Program (TCGA), are mapping the genomes of thousands of tumor types in search of mutations and expression profiles that can be targeted with new drugs.

IMMUNOTHERAPY FOR TUMOR

Personalized medicine includes several promising new methods of tumor treatment that use cells from the patient's immune system to destroy tumors (immunotherapy). The most promising methods are adoptive T-cell transfer and immunotherapy with genetically engineered T cells. Both approaches use cytotoxic T lymphocytes (i.e., T cell killers) to recognize specific antigens on the surface of malignant cells and then destroy them. Such selective tumor recognition is possible because malignant cells express many proteins specific to them that can be recognized by the immune system as foreign antigens. They appear as a result of abnormal gene expression and mutations in tumor genes. T cells are capable of recognizing tumor antigens and destroying malignant cells, but often this does not happen because tumors use many different mechanisms to suppress the T-cell response. To circumvent and overcome these mechanisms, the following personalized T-cell-based therapies have been developed.

Adoptive cell transfer. Briefly, the steps of this method can be summarized as follows:

1. Small samples containing T lymphocytes are obtained from the patient's tumor.
2. Each sample is grown in the presence of tumor material in a cell culture dish in which the conditions necessary for the cells are created.
3. As the T cells in some samples grow, they destroy the tumor cells within two to three weeks. These T cells are selected and retested for their activity against tumor cells in culture.
4. T-lymphocytes capable of destroying tumor cells are grown and multiplied under laboratory conditions to high numbers (10^{11} cells).
5. Selected T cells are injected into the patient. They are also able to divide after injection.

The process from obtaining a tumor sample to obtaining a T-cell preparation takes about six weeks.

Some clinical trials of adoptive cell transfer have yielded positive results, especially in patients with metastatic melanoma, a disease that usually has a poor outcome.

IMMUNOTHERAPY WITH GENETICALLY ENGINEERED T CELLS

Adoptive cell transfer requires activated T cells that recognize unique tumor antigens, but these are not always present, requiring alternative methods. One such method is genetically engineered T-cell therapy. It involves the creation of recombinant T-cell receptors. In other words, receptors capable of recognizing tumor antigens are created using genetic engineering.

1. T-lymphocytes obtained from the patient are injected with DNA encoding the necessary receptors. As a result, receptors are expressed on the surface of the cells.

2. Successfully altered T cells are selected, allowed to proliferate, and transfused to the patient.

Synthetic receptor genes encode either T-cell receptors structurally similar to the natural ones or chimeric antigen receptors (CARs), which are “assembled” from parts of different proteins and can directly recognize antigens on the tumor cell (activation of such T-cell by antigen-presenting cells is not required). CAR-T cell therapy has been shown to be effective in the treatment of leukemias and lymphomas in which the cancer cells are B-lymphocytes.

MOLECULAR GENETIC MARKERS OF TUMORS

Genetic diagnosis of oncological diseases. Any substance, structure, or process that can be measured to predict the occurrence of an outcome or disease can be considered biomarkers. Molecular markers used for cancer diagnosis can be divided into biochemical, immunohistochemical, and molecular genetic markers.

Changes in gene expression sequences and levels, protein structure, or function are associated with each type of tumor. Genome damage leading to cancer development can be used as quite sensitive and specific molecular genetic tumor markers. Thus, DNA- and RNA-based molecular genetic markers represent structural and functional damage detected in the genome of a tumor cell. These markers can be useful for tumor detection, monitoring the course of a disease, determining prognosis and response to treatment.

Tumor markers can be tested on tumor material, but in some cases circulating tumor cells, circulating tumor DNA, blood plasma proteins, detection of autoantibodies specific for tumor cells can be used as alternatives.

Markers can be divided into several groups, including prognostic markers, early diagnostic markers, markers that determine sensitivity to targeted medications, and others.

Prognostic markers are molecular genetic changes determined in a tumor obtained by surgical removal or biopsy: chromosomal and genomic rearrangements, mutations, deletions, translocations associated with tumor development. They allow predicting the outcome of the disease. Examples of

prognostic value molecular genetic disorders are amplification and hyperexpression of EGFR1 in bladder cancer (unfavorable prognosis), mutations in FGFR3 (favorable course). Recently, due to the development of technologies such as microarrays, it has become possible to study thousands of genes for one patient at a time.

Early detection markers are markers that allow early detection of tumors. The prognosis and treatment of a large number of tumor types depend directly on the stage at which the disease was detected. Unfortunately, detection of tumor cells through morphological analysis usually occurs at later stages, when the disease has clinical manifestations. The development of tests determining tumor-related molecular-genetic and immunohistochemical markers even at the stage of pre-tumor changes is an important direction of oncology development. A promising example of such testing is the analysis of methylation of the genome or individual genes because changes in the methylation of certain genes and their regulatory regions occur early and can be determined even before the clinical manifestation of the disease.

Markers that determine sensitivity to targeted medications help to determine an effective drug therapy for a certain type of tumor. A tumor's sensitivity to certain drugs often depends directly on the set of mutations it contains. The study of certain mutations and polymorphisms affecting the activity of enzymes involved in drug metabolism can predict the response to a certain targeted treatment and calculate the dose of medication acceptable to the patient. For example, treatment with tyrosine kinase inhibitors of epidermal growth factor receptor activity is most successful in the presence of certain mutations in the gene encoding this receptor (EGFR).

Circulating tumor DNA detected in the blood in small amounts can also be used as a diagnostic marker, which is the basis of DNA-based cancer diagnostics. Reliable tests on this basis for early diagnosis would allow early treatment to be initiated as well as increase cure rates. But for people without symptoms, most studies show low sensitivity, especially for tumors in the early stages. Nevertheless, the study of circulating tumor DNA can be used to determine prognosis (its absence after surgery is associated with a much better prognosis and lower probability of recurrence), tumor size, and disease severity (more tumor DNA in blood correlates with less favorable tumor stage or greater volume of metastases), convenient monitoring of treatment efficacy and treatment selection (sequencing of detected DNA allows choosing the therapy targeting specific mutations).

CHAPTER 10

BASIC LAWS OF INHERITANCE

BASIC CONCEPTS OF GENETICS

Genetics is the biological discipline studying heredity and variation.

Heredity is the transmission of characteristics from ancestor to descendant (from parents to children). *Variation* is the changes that occur in descendants. It causes species to evolve. The process of transferring genetic information is called inheritance.

The main patterns of inheritance were discovered by H. Mendel (1865). Traits that exclude each other (cannot exist together in the same individual) are called alternative traits. They are encoded by different variants of the same gene.

Such gene variants responsible for different alternative characters are called **alleles**. They are recorded using the same letter or combination of letters (e.g., A and a or vg⁺ and vg⁻). The source of the allele diversity of any gene is the mutations constantly happening in any population.

The sum of all traits of an individual is called *the phenotype*.

The organism's genetic makeup (i.e. the set of all genes) is called *genotype*. However, when considering the inheritance of specific traits, the genotype is the combination of alleles of the genes in question (for example, Aa or aabb).

If the alleles in the genotype of an individual are the same, it's called a *homozygote* (AA or aa), and if they are different — a *heterozygote* (Aa). In the case of the inheritance of alternative traits, a heterozygous individual cannot have both traits at once but has a trait due to only one allele.

The allele that appears in the heterozygous individual is called the *dominant allele*. The trait it determines is also called dominant.

An allele that does not phenotypically appear in heterozygotes, but only appears in homozygotes, is called *recessive*, as is its defined alternative trait.

A capital letter should be used to record the dominant allele and a lowercase letter for the recessive allele.

The crossing of individuals who differ in genotype is called *hybridization*.

Cross, in which one pair of alternative traits in parental individuals is considered, is called *monohybrid*, two pairs of traits — *dihybrid*, and more than two pairs — *polyhybrid*.

MENDELIAN INHERITANCE

The first comprehensive model explaining the basic laws of inheritance was created by Gregor Mendel based on the results of his experiments with seed pea plants. Various alternative traits (color and shape of seeds, pods, the color of flowers, etc.) were analyzed.

The experiments were conducted in several stages:

1. Plants were self-pollinated for several generations to make sure that they were pure lines (i.e., homozygotes) — in subsequent generations the analyzed trait does not change.

2. Crossing between varieties with alternative forms of the trait was carried out.

3. The hybrid offspring resulting from the crosses were self-pollinated.

Monohybrid cross. Mendel's first law is *the law of hybrid uniformity*. This law is demonstrated in an experiment in which G. Mendel crossed pure lines of pea plants with yellow and green seeds (alternative traits).

P: AA × aa

G: A a

F₁: Aa

As a result, the descendants (hybrids) in the first generation were identical in both phenotype (all plants had yellow peas) and genotype (heterozygotes).

Mendel's first law: if homozygous individuals differing in one pair of alternative characters are crossed and analyzed, uniformity of offspring individuals on phenotype and genotype is observed.

Mendel's second law — *the law of segregation*. When first-generation hybrids are crossed with each other (i.e., heterozygous individuals), the following result is obtained:

P: Aa × Aa

G: A a A a

F₁: AA Aa Aa aa

Individuals with one or two dominant alleles (AA or Aa) have yellow seed coloration, and those with two recessive alleles (aa) have green seed coloration. Consequently, the ratio of individuals according to phenotype (seed coloration) is 3 : 1 (3 parts with the dominant trait and 1 part with the recessive trait). The genotype is 1 : 2 : 1, i.e., 1 part of individuals is yellow homozygotes (AA), 2 parts are yellow heterozygotes (Aa), and 1 part is green homozygotes (aa).

Mendel's second law is formulated as follows: if first-generation hybrids (heterozygous organisms) are crossed and analyzed by one pair of alternative traits, a 3 : 1 phenotypic ratio, and a 1 : 2 : 1 genotypic ratio are observed.

Testcross. In experimental and breeding work, it is often necessary to find out the genotype of an individual with a dominant trait. For this purpose, a *testcross* is carried out — the examined individual is crossed with a recessive homozygote. If it was homozygous, then the first-generation hybrids will be uniform — all the descendants will have the dominant trait. If the individual was heterozygous, then the offspring will be expected to have a 1 : 1 phenotypic ratio.

P: aa × Aa

G: a A a

F₁: Aa aa

P: aa × AA

G: a A

F₁: Aa

Dihybrid cross. Mendel also studied the inheritance of two traits simultaneously. He used homozygous pea plants that differed in two pairs of alternative traits: the color of the seeds (yellow or green) and their shape (smooth or wrinkled). Crossing a plant with yellow smooth seeds (AABB) with a plant with green wrinkled seeds (aabb) resulted in plants with yellow smooth seeds (AaBb) in the first generation, which complied with the law of hybrid uniformity.

P: AABB × aabb

G: AB ab

F₁: AaBb

Then Mendel crossed the first generation hybrids: P(F₁) AaBb × AaBb. To analyze the results of polyhybrid cross, a *Punnett square* is usually used (Table 3).

Table 3

Results of the dihybrid cross

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

To simplify the recording of crossing results the notation “A-” means that if there is at least one dominant allele in the genotype, then regardless of the second allele the organism expresses a dominant trait.

The free combination of gametes in zygotes results in different combinations of genes. It is easy to calculate that the offspring are divided into 4 groups by phenotype: 9 parts of plants with pea yellow smooth (A-B-), 3 parts — with yellow wrinkled (A-bb), 3 parts — with green smooth (aaB-), 1 part — with green wrinkled (aabb).

Thus, a 9 : 3 : 3 : 1, or (3+1)² phenotypic ratio is observed. Thus, we can conclude that when heterozygous individuals analyzed by several pairs of alternative traits are crossed, the offspring have a (3+1)ⁿ phenotypic ratio, where n is the number of analyzed traits. The genotypic ratio is more complex but corresponds to (1+2+1)².

If we analyze the segregation for each pair of traits (yellow and green, smooth and wrinkled), we get: 9+3 yellow and 3+1 green, and 9+3 smooth and 3+1 wrinkled, i.e. phenotypic ratio is 12 : 4 or 3 : 1. Consequently, in a dihybrid cross, each pair of traits in the offspring produces segregation independent of the other pair. This is the result of the random combination of genes (and their corresponding traits), leading to new combinations of traits that were not present in the parental forms. In the example above, the original pea forms had yellow smooth and green wrinkled seeds, but in the second generation, we got plants not only with a combination of traits similar to the parent but also with new combinations — yellow wrinkled and green smooth seeds (*recombinant forms*).

Hence, Mendel's third law — *the law of independent assortment* (independent inheritance) follows: if homozygous organisms are crossed and analyzed by two (or more) pairs of alternative traits, an independent assortment of alleles of different genes and their corresponding traits is observed in the second generation.

Analyzing the results of trait segregation in the second generation (the appearance of recessive homozygotes), Mendel concluded that in the heterozygous state, “hereditary factors” do not mix, as was assumed at the time and do not change each other. Later, this notion received cytological proof — the alleles obtained from parental organisms are in different chromosomes of the same pair, and during meiosis I homologous chromosomes diverge into different cells. This concept was called *the law of purity of gametes* (W. Batson, 1902). It can be reduced to the following two main points:

- In a hybrid organism, genes do not hybridize (do not mix) but are in a pure allelic state.
- Due to the divergence of homologous chromosomes during meiosis, only one of the two alleles goes into each gamete.

Mendel's laws are statistical — observed in a group with a large number of individuals. They are also universal, i.e., the laws are typical of all eukaryotes if they reproduce sexually. However, some conditions are necessary for the appearance of Mendel's laws pattern:

- The traits analyzed must be determined by the nuclear genome rather than the mitochondrial genome.

- The genes analyzed must be not linked, i.e., they must be in different pairs of homologous chromosomes.

- There cannot be any intra- and interallelic gene interactions except for complete dominance. If they are present, the mechanism of allele inheritance remains the same, but the phenotype is determined in a more complex way.

- The probability of gametes formation of different types and survival of organisms with different genotypes must be equal (there must be no lethal alleles, Fig. 103).

- Penetrance (frequency of phenotypic manifestation) of a gene must be equal to 100 %.

- No genomic imprinting should be observed.

- Genes located in the sex chromosomes can also be inherited by a mechanism different from that described in Mendel's laws.

	A	A ^y
A	Agouti coat AA 	Yellow Coat AA ^y 
A ^y	Yellow coat AA ^y 	Dead A ^y A ^y 

Figure 103. Inheritance of the lethal allele of the agouti gene

- The phenomenon of pleiotropy, when one gene is responsible for the expression of several traits, can also cause difficulties in interpreting the results of crossing (differences from Mendel's laws).

Back and reciprocal cross. Sometimes (usually to obtain pure lines) use *backcrossing* — crossing the offspring with one of the parents (often with an individual having recessive traits). In some cases, reciprocal crossing is used. *The reciprocal crossing* is the crossing of two parental individuals (for example, AABb and aabb), in which first the maternal individual is heterozygous and the paternal individual is recessive, and then vice versa (crosses P: AABb x AABb and P: AABb x AABb).

GENE INTERACTIONS

Deviation from Mendel's laws causes various types of gene interactions (except for total dominance). Intraallelic and interallelic interactions are distinguished.

Multiple alleles. To understand intra- and interallelic interactions, it should be noted that a single gene can have more than two alleles. This is because mutations occur relatively randomly and, as a result, there can be multiple forms of almost any gene in a population. Alleles of the same gene are called *multiple* if their number in the population exceeds two. In these cases, the dominance or recessiveness of the alleles can make up a “hierarchy.”

For example, in rabbits, full dark coat color is caused by the dominant allele *A*, full white — by the recessive allele *a* (manifested in homozygotes). Complete gray coloration (chinchilla coloration) is shown in homozygotes with *ach* allele, and Himalayan coloration (the main coloration is white, and tips of ears, paws, tail, and nose are dark) — in homozygotes with allele *ah*. The allele *A* is dominant to all the others, *ach* is dominant to *ah* and *a*, and the allele *ah* is dominant only to *a*. Briefly, it can be written as follows: $A > ach > ah > a$.

Intraallelic interactions. Genetic interactions between alleles of the same gene are called *intraallelic interactions* (*allelic gene interactions*). They determine how the combination of alleles of one gene affects the phenotype.

The following types of interactions are distinguished:

- Complete dominance.
- Incomplete dominance.
- Overdominance.
- Codominance.
- Allelic exclusion.

Complete dominance. In *complete dominance*, one allele of a gene (dominant) completely suppresses the expression of another allele (recessive). In this case, homo- and heterozygotes are phenotypically indistinguishable. This interaction was observed in Mendel's experiments with snow peas. For

example, the allele for the yellow color of pea seeds completely suppresses the allele for the green color, and the allele for the *HERC2* gene that causes the development of brown eyes in humans suppresses the allele that causes the appearance of blue. Most human genetic diseases are inherited similarly. An example is phenylketonuria, a disease caused by a lack of an enzyme that metabolizes phenylalanine (phenylalanine-4-hydroxylase). Dominant homozygotes have no pathological mutations in both homologous chromosomes 12 of the *PAH* gene, while heterozygotes have a mutant gene in only one of them. Therefore, both dominant homozygotes and heterozygotes have a “normal” enzyme and have no signs of disease.

Incomplete dominance. In *incomplete dominance* (intermediate inheritance), one allele does not completely suppress the expression of the other. In hybrids of the first generation, intermediate inheritance is observed, while in the second generation, the phenotypic and genotypic ratios are the same — 1 : 2 : 1. For example, if sweet pea plants with red and white flowers are crossed, the first-generation plants will have pink flowers.

Red	White		Pink	Pink
P: AA	× aa		P (F ₁): Aa	× Aa
G: A	a		G: A a	A a
F ₁ : Aa			F ₂ : AA	Aa Aa aa
	Pink		Red	Pink White

When crossing hybrids of the first generation (with pink flowers) with each other in the second generation, we obtain the ratio of phenotype and genotype 1 : 2 : 1. Dominant homozygotes (AA) will have red flowers, heterozygotes (Aa) — pink, and recessive homozygotes (aa) — white.

This phenomenon can be explained by the dose of the gene. The dominant allele (A) determines the synthesis of red pigment, while the recessive allele (a) does not. In dominant homozygotes (AA), two genes allow the synthesis of a large amount of pigment, and the plants have brightly colored red flowers. Heterozygotes have only one active gene (A), they produce half as much pigment as dominant homozygotes, and their flowers will be lighter (pink) in color.

An example of a disease inherited in this way is familial hypercholesterolemia. In the disease, there are high levels of cholesterol in the blood, which is associated with a higher risk of cardiovascular disease. The degree of manifestation of these traits is also related to the number of abnormal alleles in the genotype.

Overdominance. In *overdominance*, the selective advantage of heterozygotes is higher than that of homozygotes for both alleles.

For example, flax plants heterozygous for a certain gene are more resistant to rust (a disease affecting these plants) than homozygotes. The mechanism of this phenomenon is that heterozygotes are immune to both races of rust, to

which different alleles determine immunity. In other similar cases, the presence of different alleles may increase the diversity of the protein subunits of the enzyme encoded by the gene, making it more active and stable.

An example of overdominance in humans is the inheritance of sickle cell anemia. The HbS allele, which causes the disease in the homozygous state, makes heterozygotes resistant to malaria, which is an advantage in environments where this disease is common.

Codominance. In codominance, the alleles of a gene are equivalent; neither suppresses the effect of the other. If they are present in the same genotype, both alleles manifest their effect. A typical example of codominance is the inheritance of human blood groups in the AB0 (AB group) and MN (MN group) systems.

Four blood groups in the AB0 system are determined by the inheritance of three alleles of one gene: I^0 , I^A , and I^B (an example of multiple alleles). The I^A and I^B alleles are codominant, and the I^0 allele is recessive to each of them. The first blood group (I, or O) is caused by the recessive I^0 allele, II (A) by the I^A allele, III (B) by the I^B allele, and IV (AB) by a combination of the I^A and I^B alleles. The gene in question encodes a special glycosyltransferase that modifies in a certain way the carbohydrate part of some glycolipids and glycoproteins on the surface of erythrocytes.

This modification differs for I^A and I^B allele products and does not occur in homozygotes for the I^0 allele. Thus, people with genotypes I^0I^0 have no A and B antigens on the surface of red blood cells, people with genotypes I^AI^0 or I^AI^A have the A antigen, those with genotype I^BI^0 or I^BI^B have the B antigen, and those with genotype I^AI^B have both antigens. Thus, the I^A and I^B alleles do not suppress each other. The presence of anti-A and anti-B antibodies is not genetically determined but develops during life due to the similarity of the described antigens to those of the human microflora (Fig. 104).

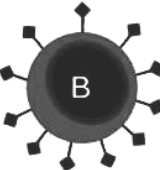
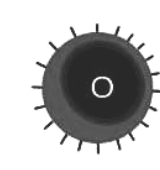
	Group A	Group B	Group AB	Group O
Red blood cell type				
Antibodies in plasma	 Anti-B	 Anti-A	None	 Anti-A and Anti-B
Antigens in red blood cell	 A antigen	 B antigen	 A and B antigens	None

Figure 104. AB0 blood group antigens

Codominance also occurs when human blood groups are inherited according to the MN system. This system is caused by the presence of two alleles, L^M and L^N . The L^M allele determines the presence of the M antigen in human erythrocytes (blood group M), while the L^N allele determines the presence of the N antigen (blood group N). The simultaneous presence of both alleles in the genotype causes the presence of both M and N antigens in erythrocytes (blood group MN).

ALLELIC EXCLUSION

Allelic exclusion, when a heterozygous organism has one allele active in some cells and another in others, also belongs to the varieties of intraallelic gene interaction. For example, in humans and mammals, each plasma cell synthesizes only one chain of immunoglobulins (recall that each of these cells produces only one “own” type of antibodies). Another example of allelic exclusion is the inactivation of one of the two X chromosomes in a female organism (described in the next chapter). The random nature of the inactivation leads to a disabling function in some cells of the maternal X chromosome and others of the paternal X chromosome.

Position effect. The *position effect* is the effect of the physical location of a gene in the genome on its expression. If a section of the chromosome changes its location (usually in mutations called translocations and inversions), normal gene expression in that chromosomal region can be altered.

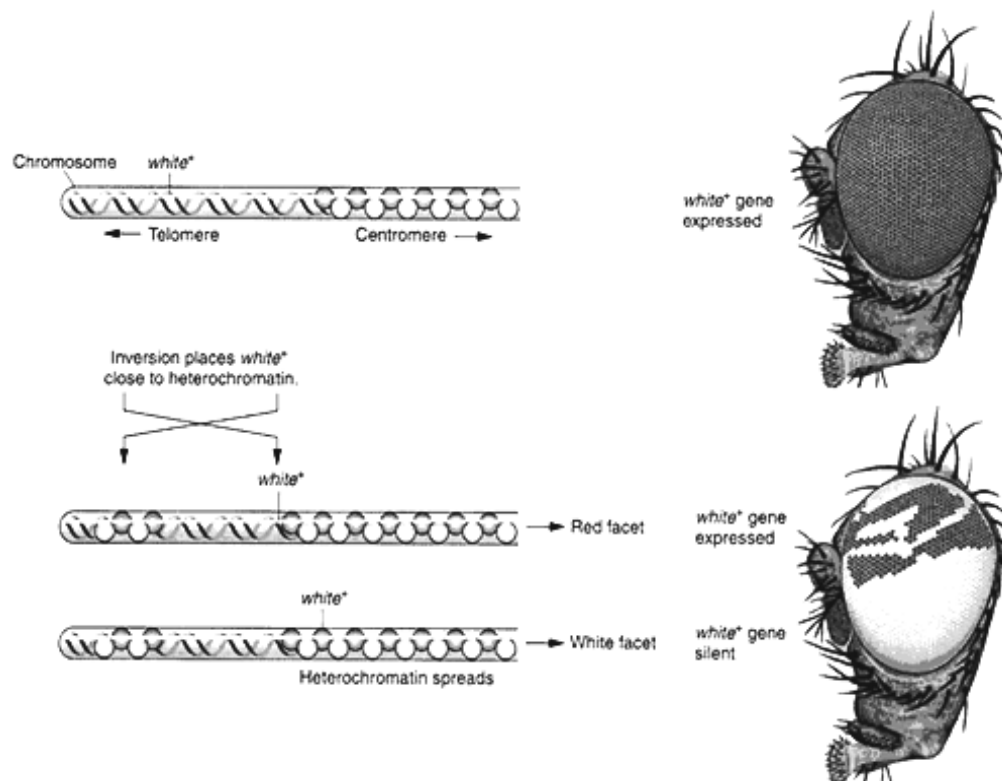


Figure 105. Position-effect variegation

For example, the expression of a gene can be suppressed if it enters a region of condensed chromatin (heterochromatin). An example of the position effect is demonstrated by female *Drosophila* flies heterozygous for the X-linked recessive white eye mutation (*w*). Flies with the *Ww* genotype usually have red eyes, but if a segment of the X chromosome containing the *W* allele translocates so that it ends up next to the heterochromatin site, the expression of the *W* allele is suppressed and the eyes become motley (Fig. 105).

Interallelic interactions. The interaction of alleles of different genes (rather than the same gene) in determining a phenotype is called an *interallelic interaction* (*non-allelic gene interaction*).

COMPLEMENTARY GENE ACTION

Complementary gene action (complementation) is an interallelic interaction in which the alleles of two different genes jointly determine the phenotypic variant of the trait manifestation. Let us consider some variants of such interaction.

1. *Two dominant alleles of different genes separately have no phenotypic manifestation, but their joint presence in the genotype results in a new variant of the trait.*

For example, crossing two sorts of sweet pea with white flowers produces hybrids with red-purple flowers:

White White
P: $AAbb \times aaBB$
G: $Ab \quad aB$
 F_1 : $AaBb$ – Red-purple flower
In the second generation it turns out:
 $P(F_1)$: $AaBb \times AaBb$
 F_2 : 9 A-B- 3 A-bb 3 aaB- 1 aabb
 Red-purple White White White
The phenotypic ratio is 9 : 7.

The mechanism of this phenomenon can be explained by the fact that pigment synthesis involves two enzymes. Each of them is encoded by the dominant allele of its own gene. In the absence of at least one of them, no pigment is produced. Thus, plants with genotype A-bb are white because of the absence of the enzyme encoded by gene B, plants with genotype aaB- are white because of the absence of the enzyme encoded by gene A, and plants with genotype aabb lack both enzymes.

A similar example is some genes in which recessive mutations lead to the development of deafness. There are about 40 such genes. For normal hearing, the dominant alleles of each gene must be present in the human genotype. If one gene is necessary for normal development of the cochlea (conventionally the D gene) and the other gene is necessary for the auditory nerve (conventionally the E gene), then people with genotype D-ee, ddE- and ddee will be deaf.

2. Alleles of one gene determine alternative variants of a character in the presence of the dominant allele of another gene, and in the presence of its recessive alleles, traits do not manifest.

An example is the inheritance of hair color in mice. Consider the two alleles of the agouti gene that determine the distribution of pigment in each hair: in the presence of the dominant allele (A), the black pigment is deposited during hair development in the form of alternating bands. In recessive homozygotes (aa), each hair is uniformly colored black (Fig. 106).

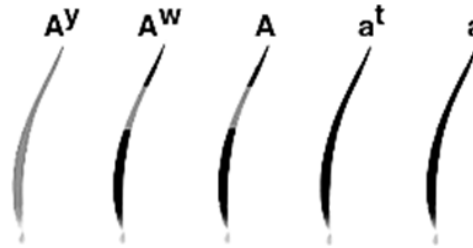


Figure 106. Effect of the five alleles of the agouti gene on coat pigmentation in mice

However, the presence of pigment requires another gene, recessive mutation of which results in white albino mice regardless of the alleles of the agouti gene:

White	Black
P: AAbb	× aaBB
G: Ab	aB
F ₁ :	AaBb – Agouti
In the second generation it turns out:	
P(F ₁): AaBb	× AaBb
F ₂ :	9 A-B- 3 A-bb 3 aaB- 1 aabb
	Agouti White Black White
The ratio is 9 : 3 : 3 : 1.	

3. Each combination of alleles of different genes has its phenotypic manifestation. This is how the comb shape in chickens is inherited (Fig. 107).

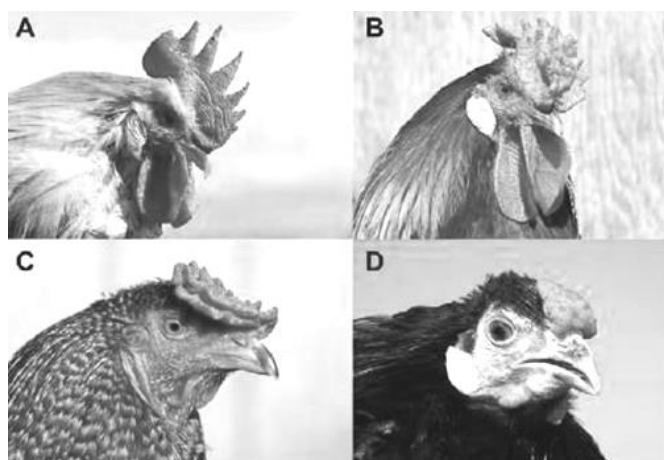


Figure 107. Four comb phenotypes in hens

Rose comb Pea comb

P: RRpp × rrPP

G: Rp rP

F₁: RrPp – Walnut comb

In the second generation it turns out:

P(F₁): RrPp × RrPp

F₂: 9 R-P- 3 R-pp 3 rrP- 1 rrpp

Walnut comb Rose comb Pea comb Single comb

Four comb phenotypes in chickens explained by segregation at the *Rose-comb* and *Pea-comb* loci and their interaction. (A) Single-combed wild-type male (*rr pp*), (B) Rose-combed male (*R- pp*), (C) Pea-combed male (*rr P-*), and (D) walnut-combed male (*R- P-*). Photos by Freyja Imsland (A-C) and David Gourichon (D).

EPISTASIS

In various contexts, the term “*epistasis*” can have two meanings. In a broad sense, epistasis refers to any gene interaction in which the phenotypic expression of one gene alters the expression of another independently inherited gene (including complementation). In a narrower sense, the suppression of the effect of one such gene by another.

Further in the text, epistasis will be regarded as an interallelic interaction in which one gene prevents the phenotypic manifestation of the other (inhibitory gene action), just as the dominant allele of a gene suppresses the effect of the recessive one.

In pigs, the dominant allele of the MC1R gene (we denote it as M) determines the black color of the coat, while the recessive (m) determines the brown color (Fig. 108).

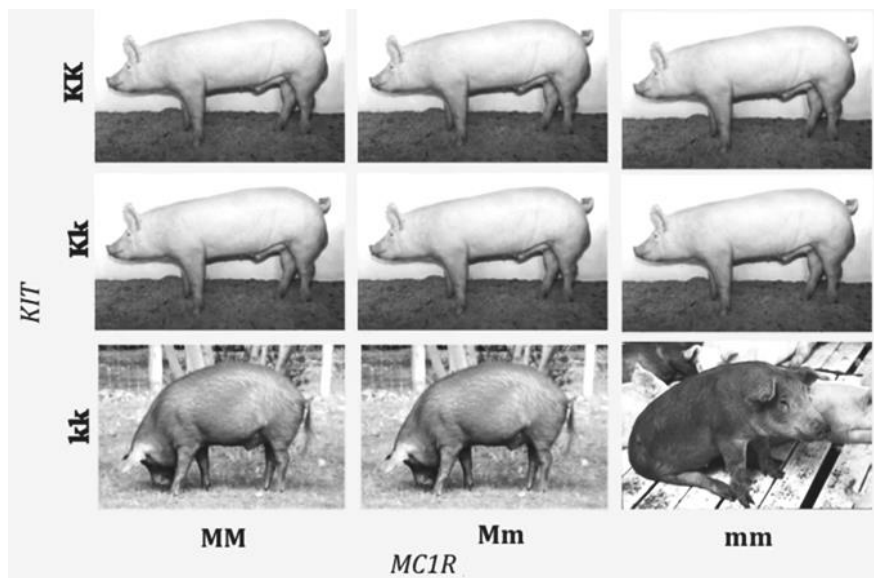


Figure 108. Inheritance of pig hair color as determined by MC1R and KIT genes

However, the dominant allele of the KIT gene (let us denote it as K) suppresses the expression of both alleles of the M gene and results in the appearance of white color. As a result, the M gene is manifested only in recessive homozygotes for the M gene (in animals with the “mm” genotype).

When crossing homozygotes with different phenotypes, all F₁ offspring have a white coat:

White Brown

P: KKMM × kkmm

G: KM km

F₁: KkMm – White

In the second generation it turns out:

P(F₁): KkMm × KkMm

F₂: 9 K-M- 3 K-mm 3 kkM- 1 kkmm

White White Black Brown

The phenotypic ratio is 12 : 3 : 1.

This example demonstrates a phenomenon called *dominant epistasis*.

Another example of dominant epistasis is the inheritance of plumage color in chickens. The dominant allele of one gene (C) determines pigment synthesis and the dominant allele of another gene (I) is its suppressor.

White White

P: CCII × ccii

G: CI ci

F₁: CcIi – White

In the second generation it turns out:

P(F₁): CcIi × CcIi

F₂: 9 C-I- 3 C-ii 3 ccI- 1 ccii

White Colored White White

The ratio is 13 : 3.

An example of *recessive epistasis* is the Bombay blood group: homozygotes for a recessive mutation of the FUT1 gene have blood group O (0), despite the presence of alleles I^A or I^B in the genotype. Only individuals containing at least one dominant (non-mutant) allele of this gene can form A or B antigens.

POLYMERIC GENE ACTION (POLYGENIC INHERITANCE)

In the previous examples, qualitative traits, whose variants are alternative to each other, were discussed. However, many traits, such as skin color and human height, are quantitative — the extent of their variation can somehow be measured. The appearance of such traits usually indicates polygenic inheritance — the number of certain alleles of different genes affects the degree of expression of a trait. Thus, according to some studies, human height is affected by variations in at least 180 genes, hair color is affected by more than 120 genes,

and human skin pigmentation is affected by more than 370 genes, many of which are involved in melanin production. In a simplified form, the genetic determination of skin pigmentation can be considered as a system of three genes in which alleles *A*, *B*, and *C*, contribute to some increase in pigmentation and alleles *a*, *b*, and *c* do not. In this model, a person with the *AABBCC* genotype would have the most pigmented skin and a person with the *aabbcc* genotype the least pigmented. The *AABbcc*, *AaBbCc*, and *aaBbCC* genotypes will have the same skin tone due to the presence of the same number of dominant alleles among the genotypes considered. The ratio in crossing triple heterozygotes is 1 : 4 : 6 : 4 : 1 (Fig. 109).

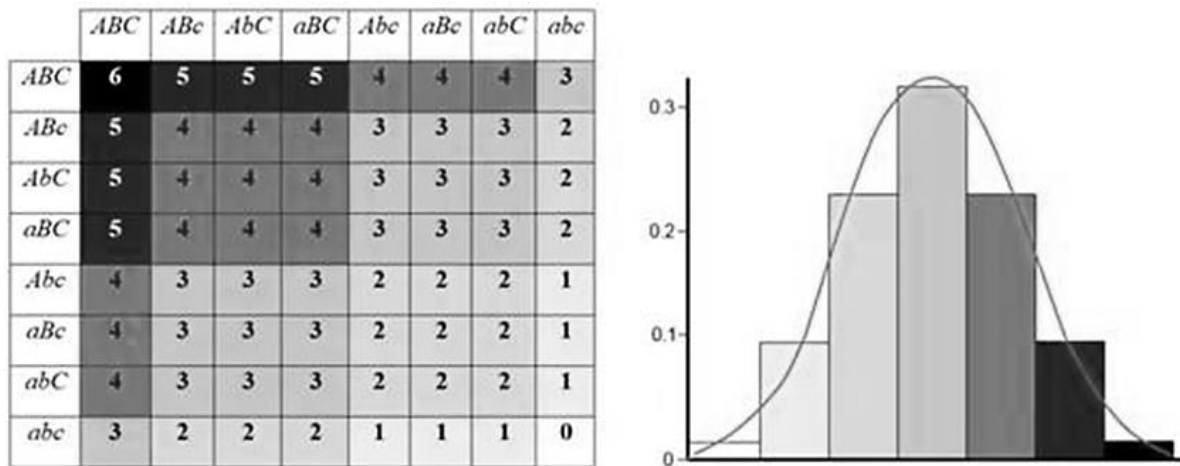


Figure 109. Punnett square and variation curve in inheriting outgrowth by a simplified three-gene system

Thus, alleles have a cumulative effect in determining the phenotype, and this kind of polymeric gene interaction is called *cumulative*.

There is also *non-cumulative* polymeric gene interaction, in which the manifestation of a qualitative trait depends on the presence of a minimum number of dominant alleles of genes in the genotype (threshold effect). The degree of trait expression does not depend on the number of dominant alleles of the polymer genes. For example, one dominant allele of any of the two genes is sufficient for the appearance of the triangular shape of the pod in the plant Shepherd's purse, and the ratio in the crossing of the double heterozygotes is 15 : 1.

Polygenic inheritance is also expressed in the presence of different risks of certain diseases such as hypertension, bronchial asthma, diabetes mellitus, schizophrenia, etc. in many people. The development of these diseases is related not only to genes, but also to environmental factors, but the combination of alleles presents at several loci in a particular person provides a greater or lesser probability of developing the disease. Because of this mechanism, such diseases are called polygenic (as well as multifactorial or diseases with hereditary predisposition).

CHAPTER 11

GENETIC LINKAGE. BIOLOGY AND GENETICS OF SEX

THE CHROMOSOMAL THEORY OF INHERITANCE

When Gregor Mendel suggested the existence of “hereditary factors” in 1860, no cellular structures in which they could be located were discovered. Around 1902, Walter Sutton and Theodore Boveri began to develop the *Chromosomal theory of inheritance*. According to this theory, genes are located at certain loci (sections) of the chromosomes. The first convincing evidence for this theory came in the early 1900s through the work of Thomas Morgan, even though he initially doubted the chromosomal theory. Morgan worked with drosophila flies (*Drosophila melanogaster*).

By crossing them, he found a male with white eyes instead of the usual red ones. The variant of the trait most commonly observed in natural populations is called the *wild type* (e.g., red eyes in *Drosophila*). Traits alternative to the wild type are called *mutant phenotypes* (e.g., white eyes in *Drosophila*).

Morgan crossed a white-eyed male with a red-eyed female. All of the F₁ offspring had red eyes, indicating that the wild-type allele is dominant. By crossing F₁ females with red-eyed males, only 50 % of the males showed white eyes, but not the females did. Morgan concluded that the gene that determines eye color in flies is located in the X chromosome, while the Y chromosome has no such gene.

GENETIC LINKAGE

Y. Setton and R. Pennett in 1908 discovered differences from the free combination of traits according to Mendel’s third law. In 1911–1912, T. Morgan et al. described *genetic linkage*, a phenomenon in which the alleles of two genes are inherited together because these genes are located in the same chromosome at a close distance from each other. Such genes are called *linked genes*.

Morgan conducted experiments on *Drosophila* flies. Alternative traits provided by two genes, *black* (B) and *vg* (V), were analyzed. Flies of the wild type had gray body color (B) and wings of normal length (V), and flies with recessive mutations of the given genes had black body color (b) and vestigial wings (v). When homozygous flies of wild type (BBVV) were crossed with homozygous mutants (bbvv), uniform hybrids having dominant traits were obtained:

P: BBVV × bbvv

G: BV bv

F₁: BbVv

Next, the reciprocal crossing was performed. Males from F₁ were crossed with recessive homozygous females and vice versa — females from F₁ were crossed with recessive homozygous males:

P: ♀ bbvv × ♂ BbVv
 G: bv BV bv
 F₁: BbVv bbvv
 50 % 50 %

P: ♀ BbVv × ♂ bbvv
 G: BV Bv bV bv
 F₁: BbVv Bbvv bbVv bbvv
 41,5 % 8,5 % 8,5 % 41,5 %

When genes were freely combined, according to Mendel's third law, the reciprocal crossing should have resulted in flies of four different phenotypes (25 % each), regardless of sex.

When F₁ males were crossed with recessive homozygotes, the progeny of only two genotypes and phenotypes identical to the parents (BbVv and bbvv) were obtained in the ratio 1 : 1. This is explained by the fact that the genes *black* and *vg* are located at a close distance in the same chromosome. As a result, they are inherited together (i.e., linked genes) as one unit of inheritance. During the first division of meiosis, one of the homologous chromosomes containing B and V alleles was transferred into one of the daughter cells, while the other one containing b and v alleles was transferred into the other.

Similar inheritance occurs when an F₁ female is crossed with a recessive homozygous male, but the female produces four rather than two types of gametes. This is because crossing over, a process in which homologous chromosomes exchange identical sections of their chromatids can occur between the given genes. In this case, not only gametes with the original combination of parental alleles (BV and bv) but also gametes with new combinations of alleles (Bv and bV) are formed. Gametes whose chromosomes contain new allele combinations relative to their parents, which is the result of crossing over, are called crossover gametes. In Morgan's experiment described above, the number of crossover Bv and bV gametes was 8.5 % each (non-crossover gametes were 41.5 % each). As a result, the number of individuals that developed from zygotes formed with the participation of crossover gametes of bbvv and bbVv (such individuals are called recombinants) was also 8.5 % each. The total percentage of crossover gametes in this experiment, i.e. 8.5 % + 8.5 % = 17 % reflects the probability of crossing over between the analyzed genes and coincides with the percentage of recombinants in the progeny.

The reason why only two types of gametes were observed in the case of the double heterozygous male is that crossing over does not occur in male *Drosophila*. Another example of such an organism is the female mulberry silkworm. One explanation for this phenomenon is that it prevents crossing over between non-homologous sex chromosomes in heterogametic sexes (see "Biology and Genetics of Sex"). Genetic linkage in which crossing over does not occur is called *complete linkage*. Genetic linkage in which crossing over is possible is called *incomplete (partial)*.

The probability of crossing over between different genes is not same: it depends on the physical distance between them. As the distance between genes increases, the probability of crossing over between them also increases. Due to

this feature, by conducting crosses and analyzing the phenotype of the offspring, one can learn such information as the order of certain genes in the chromosome and the relative distance between them. The data obtained is shown on a chromosome genetic map. The chromosome genetic map is a line segment with a schematic indication of the order of the genes and the distance between them. This distance cannot be reliably correlated with the physical distance in the DNA molecule, so it is measured in morganides (or centimorgans, cM). One centimorgan is the distance between DNA loci at which the crossing over between them occurs at a frequency of 1 %. Thus, the distance between genes in cM and the probability of crossing over between them are the same numbers. There are several reasons why the distance in cM cannot be reliably translated into base pairs. First, this distance varies among different species. For example, in yeast, it is about 2.8 kb, in *Drosophila* — about 400 kb, and in humans — about 780 kb. Secondly, the sum of the crossing over probability between pairs of genes is not always exactly equal to the crossing over probability between the outermost genes of this pair. For example, according to the experimental results, the distance between genes *black* and *vg* is equal to 17 centimorgans, but the distance between them and the gene *cn*, located between them, is 9 and 9.5 centimorgan, respectively (in total, 18.5 cM). The increasing distance between the analyzed genes increases the chance of repeated crossing over between them, which in crossing will appear as a decrease in the probability of crossing over. If the distance between genes is such that the probability of crossover exceeds 50 %, independent inheritance of genes will be observed, i.e., such genes are not considered to be linked. Therefore, the maximum observed probability of crossover is 50 %.

The actual location of individual genes is indicated on cytological chromosome maps. Such a map is a schematic representation of the chromosome showing the location of the genes. The information needed for their construction can be obtained by studying deletions — mutations in which a part of the chromosome is lost. The manifestation or non-manifestation of a recessive allele in a heterozygote with deletion is used to determine whether the deletion covers the analyzed gene.

Genes localized in one chromosome (pair of homologous chromosomes) are transmitted together (linked genes) and form one linkage group. Since alleles of the same genes are localized in homologous chromosomes, they are considered to be the same linkage group. Thus, the human genome is located in 22 autosomes and 2 types of sex chromosomes; the number of *Homo sapiens* linkage groups is 24.

Based on the data obtained, the chromosomal theory of inheritance can be stated in the following statements:

- Genes are arranged in chromosomes in linear order at specific loci; alleles of the same gene occupy the same loci of homologous chromosomes.

- Genes located in the same chromosome form a linkage group. The number of linkage groups is equivalent to a haploid set of chromosomes.
- During meiosis, crossing over — the exchange of identical chromatid segments — occurs between homologous chromosomes.
- The probability of crossing over is proportional to the distance between genes. For this reason, the distance between genes can be conventionally represented in morganides (centimorgans) — distance units equal to 1 % of the crossing over probability.

SEX, PRIMARY AND SECONDARY SEX CHARACTERISTICS

Sex is a set of morphological, physiological, biochemical, and other characteristics of an organism that determine its reproductive role. A set of characteristic morphological and physiological features of a human being that allows us to correlate them to certain sex are called sexual characteristics. They are genetically predetermined and subdivided into primary and secondary.

The primary sexual characteristics are represented by the organs directly involved in the processes of reproduction, i.e. gametogenesis and fertilization. These are the external and internal sexual organs. They are established during embryogenesis and to a large extent have already been formed by the time of birth.

Secondary sexual characteristics are not directly involved in reproduction but contribute to the attraction and meeting of individuals of the opposite sex. They develop under the influence of sex hormones and appear in organisms during puberty (about 12–15 years old). Such traits include peculiarities of musculoskeletal system development, distribution of subcutaneous fat, facial hair growth, degree of mammary gland development, voice timbre and behavior features, etc.

Somatic traits associated with sex are divided into 3 categories: sex-limited, sex-influenced and sex-linked.

The development of sex-limited traits is caused by genes located in the autosomes that are present in both sexes but are only expressed in individuals of one sex. For example, many of the genes necessary for egg-laying are present in chickens and roosters but are expressed only in chickens. In humans, an example of such a trait is the growth of a beard.

The development of sex-influenced traits is caused by genes also located in the autosomes of both sexes, but their expression and penetrance (degree and frequency of their expression) differ in individuals of different sexes. As an example, baldness in males and females. It is believed that dominant homozygotes for some genes (both females and males) do not go bald. Recessive homozygotes of males go bald earlier and females go bald later. Female heterozygotes have no hair loss, and male heterozygotes develop

baldness somewhat later than homozygotes. This situation can be interpreted as follows: the “allele of normal hair growth” is dominant in females, while the “allele of alopecia” is dominant in males.

P:	Aa	×	Aa	
G:	A	a	A	a
F ₁ :	AA		Aa	Aa
♀	normal hair growth		normal hair growth	normal hair growth
♂	normal hair growth		alopecia	alopecia

THEORIES OF SEX DETERMINATION

The development of a male or female reproductive system is controlled by genes, but sex determination is not always strictly genetically determined. In many animals, sex can be determined by environmental factors. For example, in turtles, it depends on the temperature of the environment during development. In the worm *Bonellia viridis*, the male is microscopic in size and lives permanently in the female reproductive tract. If a larva floating in water meets a suitable female on its way within 2 days, it develops into a male, and if not, into a female.

In the Atlantic herring, gender redefinition can be observed. Herring live in small shoals, each with one male and several females. If the male dies, then after a while the largest female turns into a male.

In many other organisms, including humans, sex is determined genetically at the moment of fertilization. The mechanism of this process can be varied:

- In bees, the haploid karyotype ensures the development of a male and the diploid karyotype ensures the development of a female.

- In grasshoppers and locusts, females have two sex chromosomes (XX) and males have one (X).

- In birds and butterflies, females have different sex chromosomes (ZW) and males have the same sex chromosomes (ZZ).

- In mammals, including humans, males have different sex chromosomes (XY) and females have the same sex chromosomes (XX).

The sex that has the same sex chromosomes is called homogametic because it produces one type of gametes. The sex that has different sex chromosomes is called heterogametic because it produces two types of gametes that differ in sex chromosomes. In all mammals, in *Drosophila*, the homogametic sex is female and the heterogametic sex is male.

The chromosomal theory of sex determination, which is that the sex of an organism is determined by the combination of sex chromosomes at the moment of fertilization, was suggested by C. Correns in 1907.

P:	XX	×	XY
G:	X		X Y
F ₁ :	XX		XY

The normal human male karyotype is recorded as 46, XY, and the female 46, XX.

The balance theory of sex by C. Bridges (1922) is that sex is determined by the ratio of the number of autosomes to the number of sex chromosomes. When studying the inheritance of sex in the *Drosophila* fly, it was found that one X chromosome is sufficient for the development of the male phenotype (except for some traits, for example, maturation of complete spermatozoa in such individuals does not occur). From this, it was concluded that the Y chromosome in *Drosophila* flies, like in grasshoppers and locusts, is not essential for determining the morphological sex in males. Individuals with various combinations of the number of X chromosomes and sets of autosomes (A) were obtained and their sex was studied:

2X : 2A (1 : 1) — normal females; 1X : 2A (1 : 2) — normal males;

3X : 2A (1,5 : 1) — superfemales, hypertrophied features of females, sterile;

1X : 3A (1 : 3) — supermales, hypertrophied features of males, sterile;

2X : 3A (1 : 1,5) — intersex, part of females and part of males characteristics, sterile.

Thus, the more X chromosomes in the karyotype, the more distinct the features of the female sex, and the more sets of autosomes, the stronger the features of the male sex. At an “intermediate” ratio of 1 : 1.5 (2X : 3A), signs of both sexes develop.

GONOSOMAL INHERITANCE

Traits whose development is caused by genes located in sex chromosomes are called sex-linked (X-linked and Y-linked), and the inheritance of such traits is called gonosomal or sex-linked. The X chromosome is much larger than the Y chromosome. For a long time, it was thought that the human Y chromosome, unlike the X, contained almost no genes. However, the Human Genome Project showed that the Y chromosome contains at least 75 genes (the X chromosome contains about a thousand). Some of them have homologous counterparts in the X chromosome, while others do not. Despite being related to the same pair, the X and Y chromosomes are not homologous, as they not only differ in shape and size but also have different sets of genes. Nevertheless, they do have small homologous regions, which are called pseudoautosomal regions.

They are located at both ends of the sex chromosomes, and during meiosis, there is synapsis and crossing over between them. The genes located in these regions are inherited similarly to autosomal genes. However, large parts of the X and Y chromosomes are not homologous to each other. Traits whose development is determined by genes located in a non-homologous part

of the X chromosome are called X-linked. Their inheritance, unlike the inheritance of autosomal traits, differs for representatives of different sexes, since men have only one X chromosome. Due to this fact, traits caused by X-linked genes appear in males even when they are recessive. This mechanism of inheritance has hereditary diseases hemophilia A and hemophilia B caused by mutations in genes F8 and F9 respectively, as well as color blindness caused by mutations in genes OPN1LW or OPN1MW. Such traits are transmitted from father to daughters, and from mother to daughters and sons. The presence of only one allele of a gene in the genotype is called hemizyosity.

Traits that are determined by genes located in a non-homologous part of the Y chromosome are called Y-linked or holandric. An example is hypertrichosis (excessive hair growth) of the auricles. Such traits appear phenotypically only in males and are only passed on from father to sons.

SEX DIFFERENTIATION DURING DEVELOPMENT

Sex determination is the process of determining the sex of the developing organism, which results in the development of the reproductive organs being directed toward the male or female type.

Sex differentiation is the process of developing the original undifferentiated reproductive organs into the male or female reproductive system and differentiating the sex cells into oogonia and spermatogonia. This process includes successive stages, namely: genetic, gonadal, hormonal, phenotypic, and psychological.

At the genetic stage, chromosomal sex is determined by a set of chromosomes, where XY stands for male and XX for female.

The SRY (Sex-determining Region Y) gene plays a key role in male development. It is located next to the pseudoautosomal region of the Y chromosome short arm. Its critical importance for sex determination is evidenced by several facts:

- The absence of the Y chromosome almost always results in the development of a female organism.
- Because of its proximity to the pseudoautosomal region, in rare cases, the gene is transferred to the X chromosome due to crossover errors. As a result, men may be born without the Y chromosome (de la Chapelle syndrome).
- Mutations in the SRY gene can lead to the birth of women with karyotype 46, XY (Swyer syndrome).
- In experiments using transgenic mice in which SRY is introduced into eggs that have two X chromosomes, most of the offspring develop into males.

Until the sixth week of embryogenesis, there are no morphological signs of sex differentiation in the body. Undifferentiated gonads in the future representatives of both sexes are similar in morphology and can give rise to both ovaries and testes.

At 4–5 weeks of age, the embryo develops gonads, a pair of longitudinal outgrowths of the coelomic epithelium. Each of them originates from a primary kidney (mesonephros), which functions for a short time (4th week).

The excretory duct of the primary kidney is the mesonephric (Wolffian) duct. At 6–7 weeks, the paramesonephric (Müller) duct develops next to each mesonephric duct. Later in males, the Wolffian duct differentiates into the testicular appendage, seminal duct, and seminal vesicle under the influence of androgens produced by the testicles, while the Müllerian duct regresses. In females, the Müllerian duct differentiates into the fallopian tubes, uterus, and the upper part of the vagina, while the Wolffian duct regresses.

The SRY gene becomes active in an embryo with the Y chromosome at 6–8 weeks of development. This gene encodes a protein called TDF (testis-determining factor). This protein is found in all mammals studied. TDF is a transcription factor whose action causes the transcription of genes necessary for male differentiation. Expression of the SRY gene promotes the differentiation of specialized testicular cells. This process is also promoted by the SOX9 gene, which is also regulated by the SRY gene.

At weeks 9–10 of development, SRY-induced differentiation of testosterone-producing cells occurs, which, among other things, is important for the development of the Wolffian duct in part of the male reproductive system (its absence leads to regression of the Wolffian duct). Another effect of the SRY gene is the production of a protein that causes atrophy of the Müllerian duct cells, which prevents the formation of the reproductive organs of the female reproductive system.

The X chromosome contains certain loci responsible for normal differentiation and ovarian formation, and mutations in these loci lead to the disruption of this process. The presence of only one X chromosome is also not sufficient for the normal development of the sex glands. For example, in Schereshevsky-Turner syndrome, caused by the absence of the second sex chromosome, until the 14th–16th week of development the fetus does not show abnormal ovarian development. However, in the 2nd half of pregnancy, at the stage of oocyte formation, these cells progressively atrophy and the ovaries turn into fibrous bands.

Thus, in humans, the sex of the future child is determined at the moment of fertilization depending on the combination of sex chromosomes (genetic sex). Based on the genetic sex, the gonadal sex (the presence of ovaries or testicles) develops. At puberty they start to secrete female (estrogens) or male (androgens) sex hormones intensively — the hormonal sex is formed. With the participation of the hormones, the corresponding female (ova) or male (sperms) gametes are formed — the gamete sex. At the same time, the morphological sex is finally determined — the female or male phenotype. Genetic, gonadal, hormonal, gamete and morphological sex are the physical

(morphophysiological) determinants of sex. Based on the morphophysiological sex, a corresponding entry is made in the documents (passport) — the civil sex (intermediate determinant).

There are also social and psychological determinants of gender. Boys and girls are brought up differently (gender upbringing). Based on upbringing a person develops an appropriate gender identity and gender role. Depending on one's sexual self-concept and ideas about one's sexual role, one chooses a sexual partner; in most cases, it is the opposite sex (heterosexuality), but sometimes the choice falls on representatives of one sex (homosexuality). However, as twin studies show (see the chapter "Human genetics"), such peculiarities of sexual behavior can be partially genetically determined. For example, one study identified 5 genetic variants that may be associated with same-sex sexual behavior.

SEX RATIO

The ratio of persons of different sexes varies. The sex ratios at fertilization, at birth, and at the age of childbearing are called primary, secondary, and tertiary sex ratios, respectively.

Modern data show that the primary sex ratio is close to 1 : 1, but because of the unequal probability of embryo death (more male embryos die during the first two weeks of pregnancy, and more female embryos begin to die during the second trimester), there are about 105 boys per 100 girls at birth. Around the age of 25, the number of boys and girls becomes approximately equal, and then the number of women begins to exceed the number of men, due to the higher mortality of the latter.

Inactivation of the X chromosome in the female organism. The phenotype is influenced not only by the presence of a certain gene but also by the number of its copies. The female body contains two X chromosomes and the male one, and therefore the theoretical level of production of X-linked gene products in the male and female organisms should differ by a factor of two. Since the X-chromosome can contain about a thousand genes, such differences in the level of their expression can have a negative impact on the development of the organism and, therefore, must be eliminated. In different organisms, this is achieved in different ways. For example, in

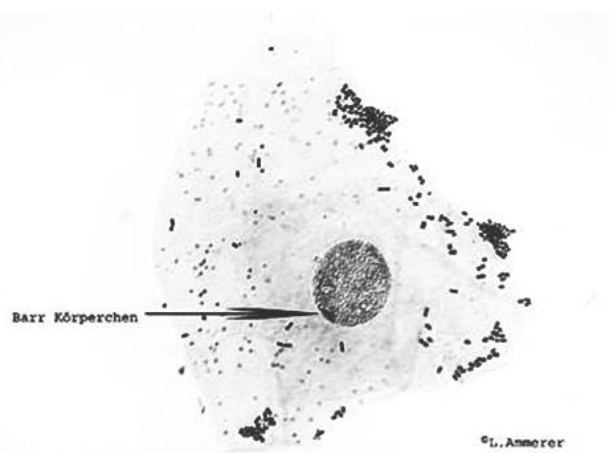


Figure 110. Barr's body in the buccal epithelium cell

male *Drosophila* there is an increase in gene expression of a single X chromosome, while in female mammals there is an inactivation of one of the X chromosomes. Such an inactivated X chromosome can be detected microscopically as a glob of intensely staining chromatin located near the nuclear membrane (Fig. 110). Such a globule was discovered in 1949 by M. Barr and was named a Barr body or sex chromatin.

Later it was established that the Barr body is a single inactivated X chromosome. It can be easily found in humans in female buccal epithelial cells (cells from the inner surface of the cheek) or fibroblasts (undifferentiated connective tissue cells).

In the early 1960s, Mary Lyon, Liana Russell, and Ernest Beutler made a suggestion that was later confirmed: Inactivation of one of the X chromosomes occurs early in embryogenesis and has a random nature — in some cells the maternal X chromosome is inactivated, while in others the paternal one. In all the descendant cells this “choice” is preserved: they have the same X chromosome inactivated as in their original progenitor cell. This explanation, called the Lyon hypothesis, was originally based on observations of female mice heterozygous for the X-linked hair color genes: the pigmentation of heterozygous females was mottled, with patches corresponding to one color allele and patches corresponding to another allele. A similar phenomenon can be observed in cats with tortoiseshell color: areas of the black coat alternate with areas of the red coat. An interesting example is one of the forms of color blindness inherited as an X-linked recessive trait. Heterozygotes can have a mosaic retina, with areas of defective color perception and surrounding areas with normal color perception.

In some chromosomal diseases, human somatic cells can have an abnormal number of X chromosomes. However, all but one of them turn out to be inactivated and are defined as Barr bodies. For example, women with a karyotype of 47,XXX have two Barr bodies. Thus, the number of these bodies is always one less than the number of X chromosomes present in the karyotype.

However, despite the activity of only one X chromosome, which is typical for healthy people, abnormalities in the number of sex chromosomes still have a phenotypic manifestation. One possible explanation is that in the earliest stages of embryonic development the X chromosomes are still active; another is that not all X chromosome genes are inactivated. Some studies have shown that up to 15 percent of human X chromosome genes avoid inactivation.

The mechanism of X-chromosome inactivation is not fully understood. Inherited by daughter cells, suppression of expression of any genes is a subject of epigenetics study. It is known that there is a region on the proximal end of the p-arm of the X chromosome that plays a key role in this process. One of its genes, XIST (X-inactive specific transcript), encodes a long non-coding RNA that is critical for X-inactivation. This RNA attracts a protein complex that

suppresses transcription at the epigenetic level. The Xist RNA spreads along the X chromosome “covering” it and leading to the formation of the Barr body. Many details of this process are currently unexplored.

Disorders of sexual development. Disorders of sexual development refer to a group of congenital conditions characterized by atypical development of chromosomal, gonadal, and/or phenotypic sex.

Disorders of sex formation can have a variety of causes.

- Changes in the set of sex chromosomes in gametes — as a result of meiosis errors, gametes may contain more than one sex chromosome or none at all. In this case, the zygote and the organism that develops from it will have an altered number of chromosomes in all somatic cells.

- Mosaicism — the cells of the organism have unequal karyotypes. Cells with an altered set of sex chromosomes can form in the early stages of embryonic development as a result of errors in mitosis. In this case, some cells will have a normal karyotype, and some will have a karyotype with either an excess or deficiency of sex chromosomes.

- Mutations in genes that affect sex determination and differentiation.

Klinefelter syndrome. Klinefelter syndrome is characterized by the presence of an extra X chromosome in the male karyotype (Fig. 111). The karyotype is 47,XXY (rarely 48,XXX or 48,XXYY), in some cases there are mosaic forms when one of the cell lines has a pathological karyotype. The Barr body is detected. The incidence is about 1 : 500 newborn boys.

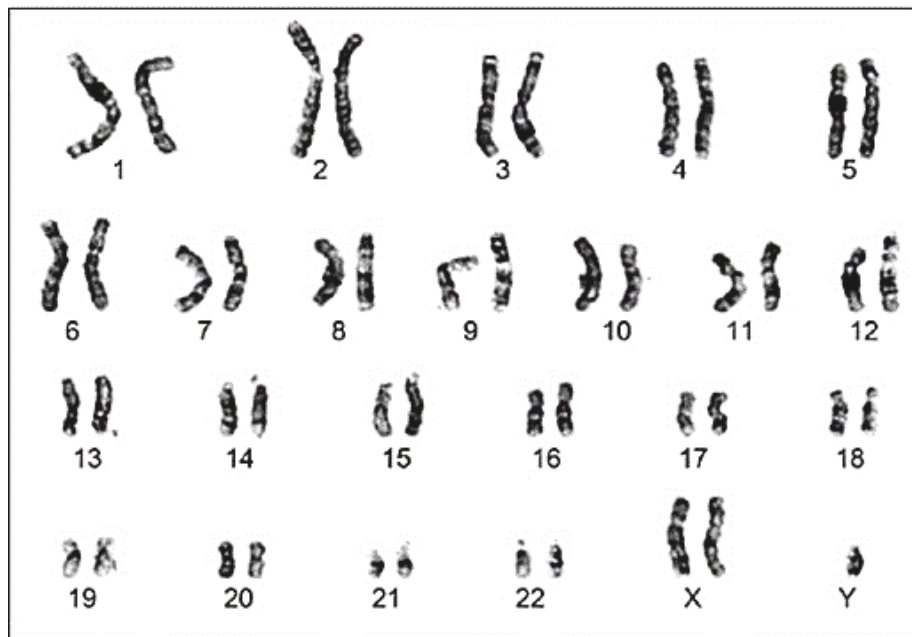


Figure 111. Klinefelter syndrome karyotype

There are abnormalities in the formation of the testes, which leads to a decrease in testosterone production. Testosterone deficiency is manifested at the beginning of puberty (symptoms are rare in childhood). High stature with

long limbs (delayed closure of the growth zone), enlarged breasts (gynecomastia), little facial and body hair, decreased fertility, and libido are observed. Spermatozoa are often absent from the ejaculate, and patients with this syndrome are infertile. Developmental delay and cognitive impairment are possible. With each additional X chromosome, there is usually a worsening of symptoms and impairment of mental abilities.

The diagnosis is made based on symptoms, and laboratory tests (characteristic changes in the concentration of testosterone and other hormones). The diagnosis is confirmed by cytogenetic studies (microscopic examination of the karyotype). Treatment includes lifelong testosterone substitution therapy.

Trisomy X. Trisomy X (triple X syndrome) is characterized by the presence of an extra X chromosome in the somatic cells of the female body: karyotype 47,XXX (Fig. 112). Two Barr bodies are detected. Sometimes there are mosaic forms (46,XX/47,XXX). The incidence is about 1 : 1000 newborn girls. This syndrome usually does not cause the appearance of significant physical features.

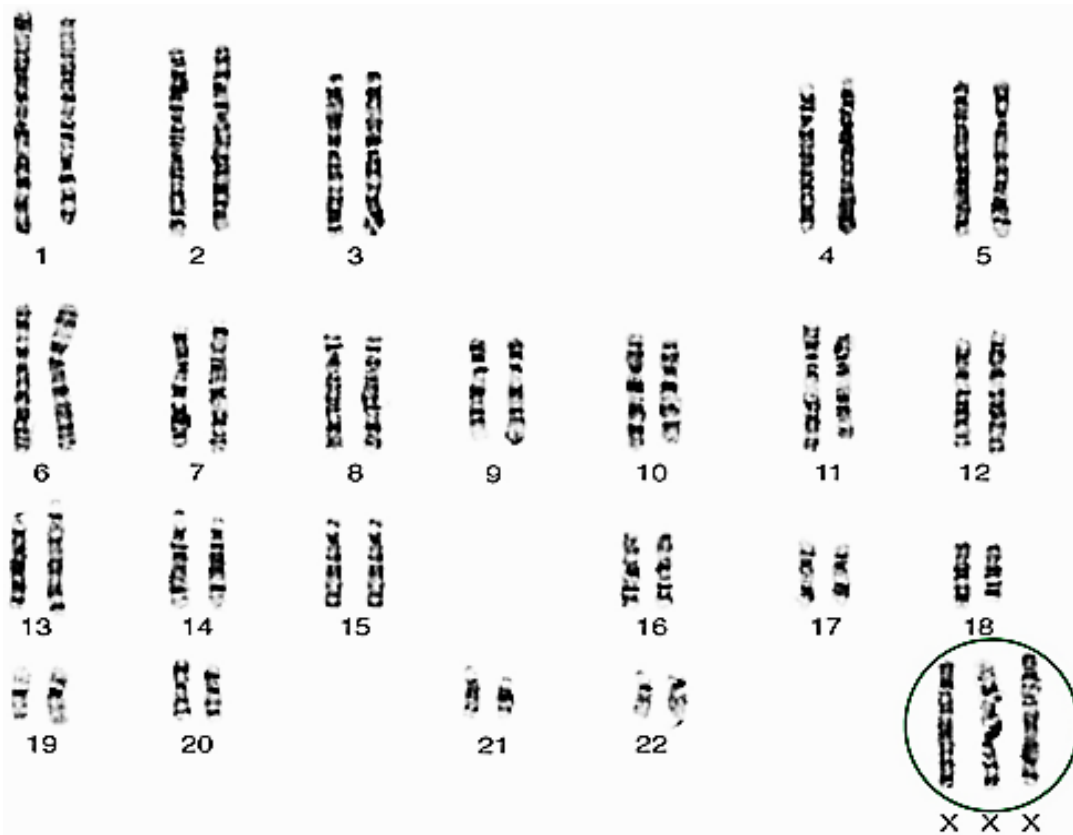


Figure 112. Trisomy X karyotype

Women with the disease may be taller than average. There is a slightly increased risk of delayed development of speech and language skills. Intellectual development is usually normal. There may be delayed development of motor skills (e.g., sitting and walking) and decreased muscle tone (hypotonia).

Sometimes there are abnormalities in the development of the kidneys. As the number of X-chromosomes increases (48,XXXX), the clinical manifestations of the syndrome increase. In most cases, sexual development is normal and fertility is not impaired, but the risk of chromosomal abnormalities in the offspring and spontaneous abortions may be increased.

The diagnosis is confirmed by cytogenetic testing.

Turner Syndrome. The syndrome was described by N. A. Shereshevsky (1925) and G. Turner (1938). Turner syndrome (X monosomy) occurs when there is only one sex chromosome (X) in the karyotype (45, X0). Barr's body is undetectable. Both complete monosomies and mosaic forms can occur. The phenotype is female. The frequency is about 1 : 2500 newborn girls. In Turner syndrome, there is low birth weight and short stature (due to the presence of only one *SHOX* gene, located on the X chromosome), ovarian development disorders, which are manifested by underdevelopment of the genital system organs, and loss of oocytes. As a result, women with this syndrome are usually infertile. In most cases, wing folds on the neck, lymphoedema of the hands and feet, chest deformities, abnormalities of the kidneys, cardiovascular system, low back hairline, and deformities of the auricles are found. Intellectual abilities are usually not impaired.

Diagnosis is based on characteristic clinical manifestations, examination of hormone levels in the blood, and confirmed by karyotype examination. Treatment includes hormone substitution therapy.

XYY syndrome. Frequency is approximately 1 : 1000 male infants. Karyotype — 47,XYY, possible mosaicism 46,XY/47,XYY

The phenotype is male and there are no specific manifestations of the syndrome, so the syndrome often remains undiagnosed. Fertility and pubertal development are normal, there is a high growth, moderate delay in motor and speech development, increased frequency, autism spectrum disorders, and learning disabilities. People with XYY syndrome are thought to have a higher level of aggression compared to karyotypically normal men with comparable levels of intelligence, but these data are not sufficiently reliable. The diagnosis is confirmed cytogenetically. Treatment is usually unnecessary.

Androgen Insensitivity Syndrome. Androgen insensitivity syndrome (AIS) is a sexual development disorder in which a person who is genetically male (karyotype 46, XY) is insensitive to male hormones (androgens). Androgen insensitivity can be either complete or partial, which determines different degrees of phenotypic expression of the syndrome. At complete androgen insensitivity, a female phenotype develops, and at a mild degree, the genitalia are formed according to the male type. Mutations in the androgen receptor (AR) gene are the cause of this syndrome.

CHAPTER 12

VARIATION. MUTAGENESIS. CARCINOGENESIS

VARIATION

Variation is the property of living organisms to acquire new traits. There are *hereditary (genotypic) variation* and *non-hereditary (phenotypic) variation*.

Hereditary variation is of two types — combinative and mutational.

– Combinative variability is the diversity of offspring of the same parents. It results from the recombination of parental genes during sexual reproduction.

– Mutational variability is the emergence of new genetic variants.

Ontogenetic variability is the change in the phenotype observed during the development of an individual.

Phenotypic variation. The *phenotypic variation (phenotypic plasticity)* is the ability of individual genotypes to produce different phenotypes when exposed to different environmental conditions. Examples in humans: variations of body weight, changes in skin pigmentation under the influence of ultraviolet, changes in the concentration of red blood cells depending on altitude, etc.

There are two main types of phenotypic variation: *morphosis* and *adaptive modification*.

Morphoses are random non-adaptive ways in which an organism changes during development. For example, some drugs taken during pregnancy may cause malformations in the fetus. Sometimes morphoses resemble *phenotypic* manifestations of a known mutation. For example, thalidomide taken by a pregnant causes serious impairments of fetal limb development, but the same effect can be caused by some mutations. In contrast to adaptive modifications, morphosis is irreversible and persists throughout the life of the organism.

Adaptive modifications are non-hereditary changes that contribute to the adaptation of the organism.

Properties of the modifications:

1. Not inherited because the genotype is not affected.
2. Reversible — modifications disappear when environmental factors change (the number of leukocytes in the blood returns to normal values at the end of the disease).
3. Adequate — the degree of trait expression is in direct dependence on the type and duration of the factor action (the longer and more intensive UV exposure, the stronger the tan).
4. The same factor usually causes the same change in individuals of the same species.
5. Directional — all individuals exposed to the same environmental factor change similarly.

6. Adaptive — useful for the organism in specific conditions (an increase in the number of red blood cells and hemoglobin content in human blood in the mountains is necessary to deliver enough oxygen to the tissues).

7. Predictable.

8. Are not material for natural selection.

9. Limited by the norm of reaction (reaction norm).

The **reaction norm** is the range within which a trait can vary. The reaction norm is genetically determined and inherited. Broad reaction norm — traits vary within a large range. As a rule, these are quantitative traits determined by multiple genes (body weight, height, skin pigmentation).

Narrow reaction norm — traits change insignificantly. Narrow reaction norms are seen in qualitative traits (hair color, eye color, and the size of a person's internal organs).

Genotypic variation. Genotypic variation is caused by changes in the DNA, which can lead to the development of new traits or a new combination of traits. It can be combinative or mutational.

Combinative variation. Combinative variation is the appearance of new combinations of traits in the progeny of the same parents. For example, if the parents have blood groups I and IV, the offspring can have either group II or III.

P: $I^0I^0 \times I^AI^B$

G: $I^0 \quad I^A \quad I^B$

F₁: $I^AI^0 \quad I^BI^0$

Mechanisms of combinative variation:

– Crossing over. Crossing over occurs at random chromosome sites.

– A random combination of chromosomes in gametes. Of the 46 human cell chromosomes, 23 are derived from the father and 23 from the mother. The gamete formed during meiosis of such a cell contains 23 chromosomes, and each of them may happen to be both maternal and paternal.

– A random combination of gametes with different genotypes (due to the mechanisms described above) during fertilization.

Combinative variation is an important mechanism of genome-wide recombination, which allows a population to “mix” alleles of different individuals, resulting in a more efficient selection of the most successful combinations.

Mutational variation. A **mutation** is a sudden, stable change in genetic material under the influence of external or internal environmental factors. Mutations are inherited.

- DNA is constantly exposed to a large number of damaging chemical reactions. There are estimates that the amount of DNA damage in one human cell ranges from 10^4 to 10^6 per day.

- In humans, mutations occur at a rate of about 10^{-6} mutations per gene per cell division, mainly due to DNA replication errors.

- Each newborn has an average of 70 new mutations that the parents did not have. More than 90 % of them are single nucleotide polymorphisms, the rests are insertions and deletions.

- The number of new mutations in each generation depends on the age of the parents: each year of the father's life adds an average of 1.5 mutations to his offspring (mother — about 0.37). The frequency of chromosomal anomalies significantly increases with the age of the mother; for example, the probability of having a child with Down syndrome for twenty-year-old mothers is about 1 : 2000, and for 45-year-old mothers — about 1 : 30.

Properties of mutations:

- Arise spontaneously.
- Inherited.
- Irreversible.
- Undirected.
- Non-adaptive.
- Inadequate for the environment.
- Individual.
- The same mutations may reoccur.
- Material for natural selection.

Mutations can be classified according to different criteria:

1. By origin (etiological factors):

- *Spontaneous mutations* — occur spontaneously.
- *Induced mutations* — produced by certain mutagenic factors.

2. By place of appearance (mutated cells):

- *Gametic (germline)* mutations occur in germ cells. They manifest in offspring.

- *Somatic* mutations occur in somatic cells. They are not inherited via sexual reproduction and manifest in the individual who has them (for example, different colors of eyes in the same person, a white strand of hair, tumors).

3. By changes in genetic material (the level where mutation occurs):

- Genome mutations.
- Chromosome mutations (chromosome aberrations).
- Gene mutations (transgenations).

Genome mutations are caused by changes in the number of chromosomes.

1. *Polyploidy* is a multiple set of chromosomes in somatic cells ($3n$, $4n$, $5n$, ...). Polyploidy is used in plant breeding and leads to higher yields. In mammals, including humans, these are lethal mutations.

2. *Haploidy* ($1n$) is a single set of chromosomes in somatic cells, such as in the drones of bees.

3. *Aneuploidy* is a decrease or increase in the number of chromosomes ($2n \pm 1$, $2n \pm 2$, etc.) that is not divisible by haploid.

Varieties of aneuploidy:

- *trisomy* — three homologous chromosomes in the karyotype, such as in Down syndrome (trisomy on chromosome 21);
- *monosomy* — one of a pair of homologous chromosomes is in the set, for example, in Turner syndrome (monosomy X).;
- *nullisomy* is the absence of a chromosome pair (lethal mutation).

Genome mutations always appear phenotypically and are easily detected by karyotyping.

Chromosome mutations (chromosome aberrations) are caused by a change in the structure of chromosomes. They can be intrachromosomal and interchromosomal.

1. *Intrachromosomal mutations* include rearrangements within a single chromosome.

- *Deletion* is the absence of part of a chromosome (partial monosomy). The deletion of a terminal chromosome region is called a deficiency. Deletions of almost any part of the chromosome can disrupt embryonic development and manifest as multiple birth defects. When the terminal regions of both chromosome arms are deleted), the chromosome often closes into a ring (ring chromosomes).

- *Duplication* is the doubling of a chromosome region. Mutations in which additional regions of chromosomes appear in the cell are called partial trisomies.

- *Inversion* is the turning of a chromosome region by 180°.

2. *Interchromosomal mutations* occur between two or more nonhomologous chromosomes.

Translocations are an exchange of segments of nonhomologous chromosomes:

- *reciprocal translocations* — two chromosomes exchange segments;
- *nonreciprocal translocations* — segments from one chromosome are transferred to the other;
- *Robertsonian translocations* — two acrocentric chromosomes are joined by their centromeric regions.

Translocations can be balanced and unbalanced translocations. Balanced translocations do not result in a loss or addition of genetic material and therefore may not have a significant effect on the phenotype. Unbalanced translocations occur when the two mutated chromosomes pass into different gametes.

Sometimes *isochromosomes* with both long or short arms are formed.

Gene mutations are associated with changes in the sequence of DNA nucleotides:

- *deletion* is the loss of one or more nucleotides;
- *insertions* are the addition of one or more nucleotides into a DNA molecule.

Such mutations change the order of codons. Let's consider the following example: deletion of the first letter in the phrase "why did the bat fly" results in "hyd idt heb atf ly". The same happens to codons in case of deletion. Such a mutation is called frameshift mutation.

Point mutations are changes in a single pair of nitrogenous bases. They include:

- *Transition* — replacement of one purine base by another purine base or one pyrimidine base by another pyrimidine base, for example, A → G, C → T; the codon in which the transition occurred is changed.
- *Transversion* — replacement of a purine base with a pyrimidine base or a pyrimidine base with a purine base, for example, A → C, G → T; the codon in which the transversion occurred changes.

Effects of point mutations:

- synonymous substitutions (*silent mutations*) do not change the amino acid due to the degeneracy of the genetic code. For example, the substitution CCC → CCT is synonymous because both CCC and CCT code for the same amino acid — proline;
- *missense mutations* — nucleotide change results in a codon that codes for a different amino acid;
- *nonsense mutations* — nucleotide change results in a stop-codon (UAA, UAG, UGA).

4. By outcome for the organism:

- *Neutral* — not affecting the fitness of the organism to its environment (most of the mutations).
- *Deleterious (negative) mutations* reduce the fitness of the organism to its environment.
- *Positive* — increasing fitness of the organism to its environment. Positive mutations are rare but are important for evolution.

5. According to the type of allelic interactions:

- Dominant;
- Recessive.

CAUSES OF MUTATIONS

Most mutations arise spontaneously without visible environmental influences during spontaneous hydrolysis of chemical bonds, replication errors, etc. They can also be caused by mutagenic influences of different natures.

Replication errors. Replication errors arise because the accuracy of DNA polymerase functioning is not absolute. Sometimes they insert non-complementary nucleotides into DNA during replication. DNA polymerases are able to correct most of these errors, but this does not always happen.

In addition, small insertions or deletions also can occur during replication, especially in the regions containing tandem repeats.

Depurination is the breaking of the glycosidic bond between deoxyribose and a purine nitrogenous base in a nucleotide. If the apurinic sites are not repaired, DNA polymerase can introduce a random nucleotide into that site during replication. Loss of pyrimidine bases also occurs in the cell, but its frequency is many times less.

Deamination, along with depurination, is one of the most common causes of spontaneous mutations. During deamination, the amino group in cytosine or adenine is converted into a keto group. Cytosine is converted into uracil, and adenine is converted into hypoxanthine (a guanine-like compound). Thus, after the conversion of C to U, the initial GC pair is converted into an AU pair, and then, during the next replication, into AT.

Mechanisms of chromosomal and genomic mutations. The occurrence of deletions and duplications is explained by the phenomenon of unequal crossing over. Normally, homologous chromosomes exchange identical regions. Unequal crossing over is an error in which non-identical regions are exchanged. This results in a duplication in one chromosome and a deletion in the other one.

Translocations are associated with double-stranded breaks in DNA (occurring at a rate of 10 to 50 per cell per day). Due to an error in the repair of the break, the end of one chromosome is mistakenly joined to the end of another one.

Polyploidy in humans can occur when more than one sperm cell enters the ovocyte during fertilization (2–3 % of pregnancies).

Aneuploidy is based on errors in attaching the microtubules of the fission spindle to the chromosome. As a result, phenomena such as anaphase lag, sister chromatids, or homologous chromosomes may occur during mitosis or meiosis.

Stability of genetic material to damage. The stability of genetic material to damage is ensured by:

- DNA double helix;
- diploid set of chromosomes;
- degeneracy (redundancy) of the genetic code;
- repeats of some genes;
- DNA repair.

Antimutagens significantly reduce the frequency of mutations. They include about 200 compounds:

- some amino acids;
- some vitamins;
- pharmacological agents.

DNA REPAIR

DNA repair is an intracellular process that provides restoration of the normal base-pair sequence and structure of damaged DNA. DNA damage may include chemical modifications of nitrogenous bases, breaks of one or both DNA strands, insertions or deletions of nucleotides, and cross-links of adjacent pyrimidines (for example, T-T).

Proofreading. Sometimes DNA polymerase adds a wrong nucleotide into the new DNA strand. But the enzyme corrects its own mistakes by removing the wrong nucleotide and adding the correct one. This process is called *proofreading*.

Direct reversal. Direct reversal is the removal of the damage from nucleotides without replacing them. For example, if a methyl group (-CH₃) is added to a certain guanine (G) region, the enzyme methyltransferase will remove it. Another example of direct reversal is the repair of T-T dimers by the enzyme photolyase, which destroys the dimers. Photolyase is an ancient enzyme that is present in many organisms, but not in humans.

REPAIR WITH EXCISION

Base excision repair detects and removes damaged nitrogenous bases. For example, cytosine turned into uracil by deamination is removed by the enzyme glycosylase. After the base has been removed, the rest of the nucleotide is also removed, and DNA polymerase fills the gap with a new nucleotide, and then DNA ligase connects the nucleotides.

Nucleotide excision repair detects and corrects damage that changes the shape of the DNA double helix.

For example, UV radiation, two adjacent pyrimidines can bind to one another. This changes the shape of the double helix and causes errors in DNA replication. The most common example of such mutation is the thymine dimer.

In excision nucleotide repair, damaged nucleotides are removed along with adjacent nucleotides:

1. DNA damage that causes double helix deformation, such as a thymine dimer, is first recognized by a special protein complex.
2. This complex attracts a helicase, which unwinds the double helix.
3. Specific proteins bind to the complex and cut the damaged DNA strand at points 24–32 bases apart on each side of the damage.
4. A single-stranded DNA fragment with damaged bases is released and degraded to mononucleotides.
5. The gap is filled in with new nucleotides by DNA polymerase, and the parts of the strand are joined together by DNA ligase.

Mismatch repair eliminates base-pair mismatches and insertions or deletions of one or more nucleotides that are accidentally introduced by DNA polymerases during replication. Mechanism of DNA mismatch repair in human cells:

- Special proteins bind to the damaged DNA sites and attract an endonuclease, which cuts the newly synthesized strand.
- Then, helicase unwinds the double helix of DNA, and an exonuclease removes some nucleotides, including the non-complementary one.
- Then, as in base excision repair, the gap is filled with nucleotides by DNA polymerase, after which the fragments are joined by DNA ligase.

REPAIR OF DOUBLE-STRAND BREAKS

There are two ways to repair double-stranded DNA breaks: nonhomologous end joining and homology-directed repair.

In the case of **nonhomologous end joining**, the ends of the DNA are joined. This mechanism is the most common way to repair double-strand breaks. It is prone to errors and often leads to the loss (or insertion) of several base pairs at the junction site. As a result, a such repair can cause mutations.

Homology-directed repair uses the same region of a homologous chromosome or a sister chromatid as an example to repair the break.

Mutations in the genes involved in DNA repair are known to cause many disorders. For example:

- Lynch syndrome is caused by mutations in genes necessary for mismatch repair. Tissues of sick people accumulate mutations much faster than healthy individuals. This can lead to the development of colorectal cancer.
- Xeroderma pigmentosum (XP) is caused by mutations in the genes necessary for nucleotide excision repair. As a result, thymine dimers and other types of UV-induced are not repaired. People with xeroderma pigmentosum suffer from freckles, capillary dilation, keratinization of the epidermis, eye damage, and severe sunburns when exposed to sunlight. About half of them get skin cancer before the age of 10 unless they avoid the sun's rays.
- In Fanconi anemia, there is a lack of bone marrow function, leading to a decrease in all blood cells and hyperpigmentation. This disease is associated with mutations in the genes encoding proteins that participate in DNA repair and replication as well as other cellular processes.

CARCINOGENESIS, THE CONCEPT OF ONCOGENES AND TUMOR SUPPRESSOR GENES

Carcinogenesis (oncogenesis) is the process of tumor formation from a normal cell. Tumor — an abnormal mass of tissue that develops as a result of uncontrolled cell proliferation (division). Tumors can be benign or malignant.

Benign tumors may increase in size but not spread to other tissues and parts of the body. Malignant tumors can spread to other tissues and can spread to other regions of the body through the blood and lymphatic systems (metastasis). The branch of medicine that deals with the prevention, diagnosis, and treatment of tumors is called oncology.

There are several basic features of cancer cells, among them:

- Ability to sustain chronic proliferation.
- Resistance of apoptosis.
- Ability to have an unlimited number of divisions.

Cancer has a genetic nature and is caused by malfunctions of genes regulating cell division (proteins that directly stimulate progression through the cell cycle, transcription factors necessary for their production, growth factors, and their receptors, proteins involved in signal transduction from the receptor to the transcription factors). Mutations in the genes encoding such proteins can lead to uncontrolled cell division.

Two groups of genes contribute to tumor development: *proto-oncogenes* and *tumor suppressor genes*. For a better understanding of their function, they are sometimes compared to the gas and brake pedals in a car, respectively.

Proto-oncogenes encode proteins that ultimately promote cell division. Thus, the transformation of a normal cell into a malignant cell is usually associated with the overexpression of proto-oncogenes. If a proto-oncogene is abnormally expressed and contributes to tumor development, it is called an oncogene. Examples of proto-oncogenes are the genes of the ras family, mutations of which are found in 30 % of human tumors.

Tumor suppressor genes are genes whose products stop progression through the cell cycle in response to DNA damage or external signals that inhibit proliferation. They usually regulate cell cycle checkpoints or initiate the process of apoptosis. Suppression of the expression of such genes promotes tumor development. An example is the TP53 gene, which is mutated in over 50 % of all cancers.

The most common mechanisms of proto-oncogene activation:

1. Activation under the action of viral genes embedded in the genome.
2. Activation by translocation (for example, to a site regulated by another promoter).
3. Activation by amplification (copy multiplication) of the proto-oncogene.
4. Activation by proto-oncogene mutations.

Not one but several mutations must occur to turn a normal cell into a malignant one.

CHAPTER 13

POPULATION GENETICS

GENERAL CHARACTERISTICS OF POPULATIONS

A **population** is a group of individuals of the same species occupying a defined geographic area, freely crossing within the group, and relatively isolated from the same groups of a given species. Speaking of human populations, they can be defined as all the inhabitants of a country or an island or all the inhabitants of a particular small town or village. The genetic information carried by members of a population constitutes that population's **gene pool**. The gene pools of populations form the gene pool of a species. Individuals of the same population have different genotypes, i.e. a population has **genetic polymorphism**. The key laws of population genetics allow us to calculate the frequencies of alleles and genotypes in a population's gene pool.

HARDY–WEINBERG EQUILIBRIUM

The relationship between allele frequencies and frequencies of different genotypes in the gene pool is described in a mathematical model developed by Godfrey Hardy and Wilhelm Weinberg. This model is called the **Hardy–Weinberg equilibrium (HWE)**, also called the Hardy-Weinberg principle. It states that in an ideal population, the ratios of gene and genotype frequencies are constant and do not change from one generation to the next. An ideal population refers to an infinitely large population with random mating (panmixia) and no mutations, natural selection, or migration.

The frequency of alleles in the population is calculated using the formula:

$$p + q = 1,$$

where “p” is the frequency of the dominant allele and “q” is the frequency of the recessive allele.

For example, if the frequency of allele A in the population is 70 % (i.e. $p = 0.7$), the frequency of allele a should be $q = 1 - 0.7 = 0.3$ (i.e. 30 %).

Note that the previous formula does not calculate the frequencies of individuals with different genotypes, because each individual has two homologous chromosomes containing the analyzed gene. Genotype frequencies are calculated using the formula:

$$p^2 + 2pq + q^2 = 1,$$

where “ p^2 ” is the frequency of dominant homozygotes (genotype AA), “ $2pq$ ” is the frequency of heterozygotes (genotype Aa), and “ q^2 ” is the frequency of recessive homozygotes (genotype aa).

Consider this with the example of a population where $p = 0.7$, and $q = 0.3$. If the frequency of dominant homozygotes is calculated as p^2 , it should be $0.7^2 = 0.49$. Similarly, the frequencies of heterozygotes and recessive homozygotes can also be calculated. They are $2pq = 0.42$ and $q^2 = 0.09$. Let's

substitute these values into the formula $p^2 + 2pq + q^2 = 1$. Thus, $0.49 + 0.42 + 0.09 = 1$, i.e. the obtained values correspond to the formula.

Let's try to prove that these frequencies will not change in the next generation. Consider all possible genotypes and their frequencies. Since we are considering autosomal genes, they will be equal for both males and females:

$$\text{♂ } 0.49(AA) + 0.42(Aa) + 0.09(aa)$$

$$\text{♀ } 0.49(AA) + 0.42(Aa) + 0.09(aa)$$

Thus, the chance that a randomly selected pair from the population will consist of two dominant homozygotes ($\text{♂}AA \times \text{♀}AA$) is $0.49 * 0.49 = 0.2401$. All their children will have genotype AA. In other words, 24.01 % of families in this population will have children with the AA genotype. The chance of pair formation $\text{♂}Aa \times \text{♀}AA$ is $0.42 * 0.49 = 0.2058$. In such families, children with genotypes AA or Aa can be born, i.e. dominant homozygotes born in such families will make $20,58 \% / 2 = 10,29 \%$ of all children in the population. The same situation is observed for pairs $\text{♂}AA \times \text{♀}Aa$. In addition, children with the AA genotype can be born in $\text{♂}Aa \times \text{♀}Aa$ families: $\frac{1}{4}$ of 0.422, i.e. 4.41 % of children in this population. Let's calculate the frequency of AA genotypes in the new generation: $0.2401 + 0.1029 + 0.1029 + 0.0441 = 0.49$, i.e. $p^2 = 0.49$, as in the previous generation. The same calculations can be performed for heterozygotes and recessive homozygotes. They will be the same as those in the "parent generation", so we can conclude that in this population allele and genotype frequencies do not change with the generations, which is confirmed by the Hardy–Weinberg principle.

Application of Hardy–Weinberg equilibrium. The Hardy–Weinberg equilibrium can be applied to human populations. For example, it is known that the incidence of phenylketonuria, an autosomal recessive disease caused by mutations in the *PAH* gene, is about 1 : 6000 in the Republic of Belarus. This ratio reflects the frequency of recessive homozygotes, i.e. $q^2 = 1/6000 = 1.67 * 10^{-4}$. To calculate the frequency of the mutant allele we extract the square root of this number and get $q = 0.0129$. Hence, $p = 1 - q = 0.9871$. Based on these data we can calculate the approximate frequency of phenylketonuria carriers in Belarus: $2pq = 2 * 0.9871 * 0.0129 = 0.025$. Thus, we can conclude that about 2.5 % of Belarusians are carriers of mutations in the *PAH* gene, which is 150 times higher than the frequency of the disease itself.

CALCULATING FREQUENCIES OF X-LINKED ALLELES

In mammals, females have the XX genotype, and males have the XY genotype. Therefore, in human populations, the frequency of a certain X-linked allele in the gene pool is equal to the frequency of males having the trait caused by this X-linked allele. In other words, since men have only one X chromosome, the probability to inherit an X chromosome with the given allele must be equal to the frequency of this allele. For women, the previously used formula would

be true, and the probability of being homozygotes would be equal to q^2 , where q is the frequency of the allele in question.

Let's take the example of X-linked recessive color blindness. It affects about 8 % of men. In this case, the frequency of the color-blindness allele is $q = 0.08$. In other words, 8 % of X-chromosomes in the population carry this allele and the remaining 92 % carry the dominant allele responsible for normal red-green color vision ($p = 0.92$).

Therefore, the frequency of this form of color blindness among females is $q^2 = 0.08^2 = 0.0064$ and the frequency of female carriers is $2pq = 2 * 0.08 * 0.92 = 0.147$, or 14.7 %.

EQUILIBRIUM IN A POPULATION

Recall that the Hardy–Weinberg principle is a mathematical model that describes a non-existent ideal population. In real human populations, however, variations from the expected allele and genotype frequencies may occur. These variations will be particularly pronounced as generations pass and as the population size decreases. One way to check if the Hardy–Weinberg assumptions in a given population are not violated is to determine if the population genotypes are in equilibrium. To do this, we first determine the genotype frequencies, and from the genotype frequencies, we can calculate the allele frequencies. Then, the allele frequencies in the parental generation are used to predict the genotype frequencies in the next generation. According to the Hardy-Weinberg principle, genotype frequencies should correspond to the ratio $p^2 + 2pq + q^2 = 1$. If this is not the case, it means that one or more assumptions are invalid for this population.

For example, let us imagine a population in which the genotypes of 200 individuals were determined using molecular-genetic methods. 33 individuals had the genotype AA, 96 individuals — Aa, and 71 individuals — aa. In this case, the frequencies of genotypes are $p^2 = 33/200 = 0.165$, $2pq = 96/200 = 0.48$, $q^2 = 0.355$. Based on these data we can calculate the allele frequencies. Each of the 200 people has two alleles, so there are 400 alleles in total. Let's calculate the proportion of dominant alleles (present twice in homozygotes (AA) and once in heterozygotes — Aa): $p = (33 * 2 + 96)/400 = 0.405$. Similarly, let's calculate the frequency of recessive allele: $q = (71 * 2 + 96)/400 = 0.595$.

Based on the obtained data on alleles, let's predict the frequencies of genotypes and compare them with the real frequencies obtained from the study of 200 people:

$p^2 = 0.405^2 = 0.164$, which is close to the value in a sample of 200 people (0.165)

$q^2 = 0.595^2 = 0.354$, which is also close to the value in the sample of 200 people (0.355) $2pq = 2 * 0.405 * 0.595 = 0.482$ in the sample of 200 people — 0.48.

Thus, the real genotype frequencies practically matched the predicted ones, and we can conclude that the population is in Hardy–Weinberg equilibrium. However, the differences in such calculations could be large, and estimating their significance “by eye” can be problematic. In that case, statistical methods can be applied to estimate their significance.

Factors that change allele and genotype frequencies in a population.

As mentioned earlier, the Hardy–Weinberg principle is a mathematical model describing an ideal population that does not exist in nature. In real populations the number of individuals is limited, mating is not random, there are mutations, and natural selection and migration are possible. In nature, these forces — natural selection, mutations, nonrandom mating, migration, and genetic drift — change frequencies of alleles and genotypes in the population’s gene pool (microevolution) and cause the formation of new taxa (macroevolution).

Mutations as a source of new alleles in the gene pool. In each generation, the alleles of the population combine randomly, producing new genotypes and phenotypes. However, new alleles appear due to mutations. According to various estimates, in human populations, each newborn has from several dozen to more than a hundred new mutations. Most of them originate as DNA replication errors in precursors of gametes. Mutations occur randomly, without regard to possible benefits or harm to the organism.

In addition to creating new alleles, mutations can change the frequencies of existing alleles, which is especially noticeable in small populations.

However, the probability of such an event depends on the mutation frequency of the gene. For example, new mutations in the gene *FGFR3* causing achondroplasia (a disease manifested by dwarfism due to the underdevelopment of long bones) occur in 1.4 : 100 000 people. Thus, in large populations, the effect of mutations on allele frequencies is small.

NATURAL SELECTION

The most important mechanism that changes the frequencies of alleles and genotypes in populations is natural selection.

Natural selection is the process by which the number of organisms better adapted to their environment increases with every generation as they produce more viable and fertile offspring than other organisms.

In population genetics, natural selection is the process by which the proportion of an allele increases in the next generation. This may be due to the effect of a given allele on an organism’s survival, fertility, behavior, and other factors.

Imagine an infinitely large ideal population in which the frequency of the dominant allele *A* is 1 % ($p = 0.01$). In our example, individuals with the genotype *aa* (not having the allele *A*) on average leave 5 % fewer offspring than individuals having this allele (genotypes *Aa* and *AA*). In this case, the frequency of allele *A* will increase approximately by 5 % with each generation. After 200 generations it will be close to 80 %, and after 700 generations — near

95 %. In this mathematical model, 100 % is unattainable due to the infinite size of the population, but in real populations with limited numbers, the 100 % frequency will be reached, and the allele *a* will be completely displaced (Fig. 113).

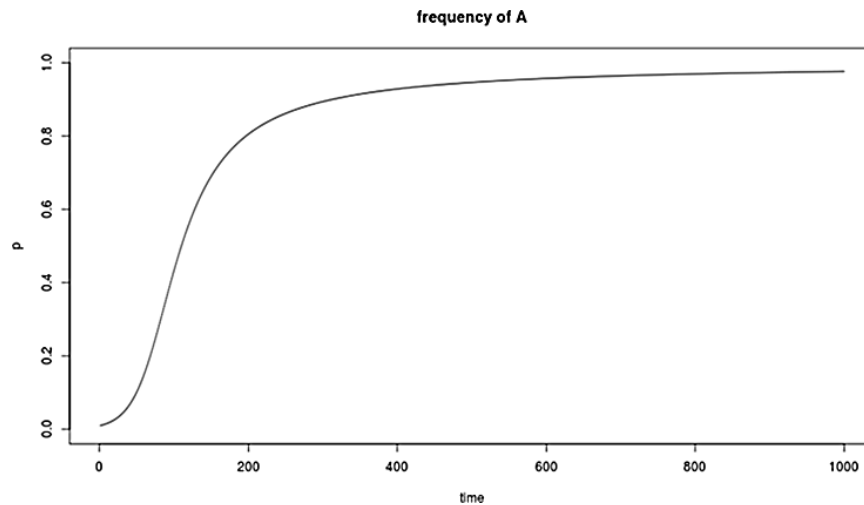


Figure 113. Modeling of natural selection in an infinitely large population for 1000 generations. X represents generations, and Y represents *p* (frequency of dominant allele). Individuals with genotype *aa* leave 5 % less offspring than individuals with genotypes *AA* and *Aa*, with an initial value of *p* = 0.01. The resource <https://phytools.shinyapps.io/selection/> was used for modeling

The genetic contribution of an individual organism to future generations is its **fitness**. Genotypes with a higher average level of reproductive success have high fitness, while genotypes with a low level of reproductive success have low fitness. The letter *w* is used to quantify it. Thus, in the example above, $w_{AA} = 1$ and $w_{Aa} = 1$, but $w_{aa} = 0.95$.

The considered example shows a positive selection for alleles that increase fitness. The allele *a* can be considered a deleterious allele as it decreases fitness. Negative selection eliminates deleterious alleles, such as alleles that cause hereditary disorders. Fitness depends on the environment of the population. For example, the skin pigmentation of *Homo sapiens* evolved differently depending on the environment. Humans in ancestral populations were dark-skinned. As human populations began to migrate, the northern populations accumulated mutations decreasing pigmentation. This is explained by the fact that lowering the chance of vitamin D deficiency (one of the steps of its synthesis occurs in the skin under the influence of ultraviolet (UV) light). In regions closer to the equator, where the level of UV radiation is much higher, such mutations were deleterious, but the genetic variants increasing skin pigmentation, provided better protection against UV.

In some cases, the frequency of alleles that look deleterious may increase as they do not decrease, but increase fitness. For example, the relatively high frequency of cystic fibrosis (caused by mutations in the *CFTR* gene) in European countries can be explained by the fact that in the past, when diseases such as cholera were common, carriers of the disease may have been more

resistant to fluid and electrolyte loss. In the current environment, where the relevance of such diseases has become much lower, the deleterious effect of such mutation on the population outweighs the benefit. The high prevalence of sickle cell anemia in many African countries has a similar explanation.

GENETIC DRIFT

Previously, we considered the evidence that the frequencies of alleles and genotypes in a population do not change in subsequent generations. However, we used an ideal mathematical model in which individuals with all genotypes left an equal number of offspring. In real populations, this is not always the case: due to random factors, some individuals may leave more or fewer offspring, the number of children in families is not equal, etc. As a consequence, in natural populations, the frequencies of alleles and genotypes in subsequent generations are similar, but not equal to those in previous generations. This effect is called **genetic drift** — random fluctuations in the frequencies of genes and genotypes in a population over generations. Genetic drift is especially strong in small populations (Fig. 114).

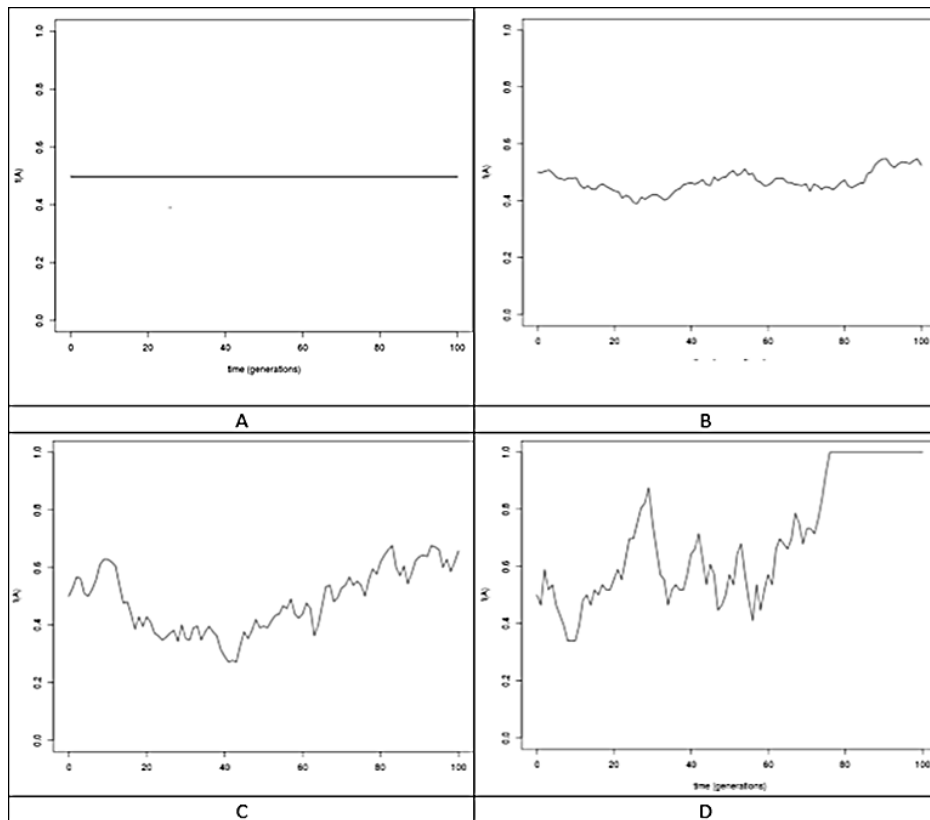


Figure 114. Modeling of the frequency of dominant allele A with an initial frequency of 0.5 in different populations during 100 generations. X-axis is generations, Y-axis is p (frequency of dominant allele). A — the ideal population. No genetic drift is observed due to infinite numbers. B — population of 1000 individuals. Genetic drift is observed, but the average allele frequency remains close to 50 %. C — population of 100 individuals. An increase in genetic drift is observed. D — population of 30 individuals. Strong drift is observed, leading to the complete elimination of the recessive allele and fixation of the dominant allele by the 80th generation

If the population size decreases, the chance that the frequency of one of the alleles during random fluctuations will reach 100 % increases significantly. This should inevitably happen in any population, but in small populations, the process is much faster.

Genetic drift can also affect natural selection. In small populations, the effect of drift on allele frequencies can become stronger than the effect of natural selection. As a result, positive alleles may be lost, and negative mutations may reach 100 % frequency (Fig. 115).

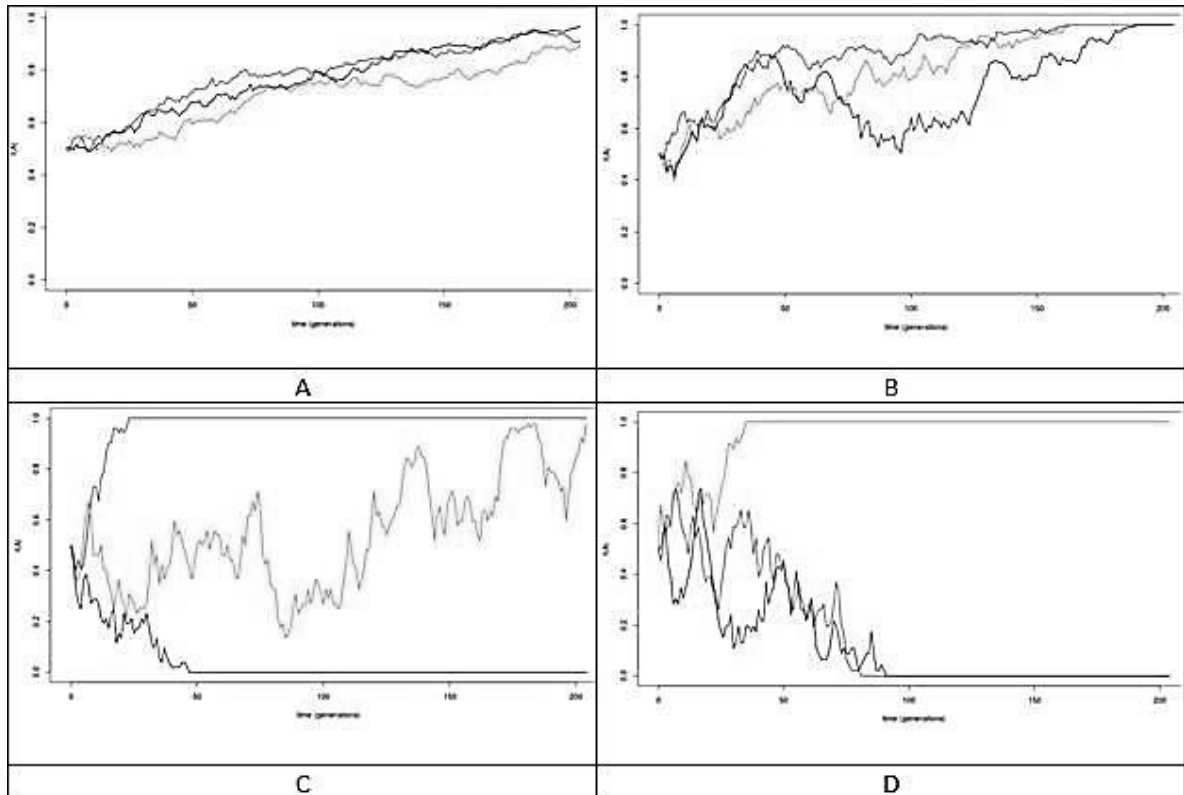


Figure 115. Modeling of natural selection for the dominant allele A with its initial frequency of 0.5 in different populations for 200 generations. Individuals not having this allele leave 5 % fewer descendants. On the X-axis — generations, on the Y-axis — p (frequency of the dominant allele). A — model for three populations of 1000 individuals. B — the model for three populations of 200 individuals. C, D is a model for three populations of 25 individuals. Genetic drift is so strong that in some cases, contrary to natural selection, the frequency of allele A reaches 0 %, while the more deleterious allele a reaches 100 %. The resource used for modeling was <https://phytools.shinyapps.io/drift-selection>

In addition to small population size, drift can occur because of the **founder effect**, which occurs when a population comes from a small number of individuals (Fig. 116). Imagine a population where the incidence of a genetic disorder is 1 : 10,000. Using the calculations described earlier in this chapter, we can determine that the heterozygote frequency would be 1 : 100. Imagine 100 members of the population establish a new settlement, and by chance 3 of them are carriers. Then the incidence of the disorder in the new

population may approach 1 : 1000 even if its number increases by orders of magnitude. A similar example of genetic drift is the **bottleneck effect**. It is observed in cases where the population size sharply decreases. Then there is a recovery in abundance, but due to increased drift in small groups, allele frequencies can change dramatically (Fig. 117).

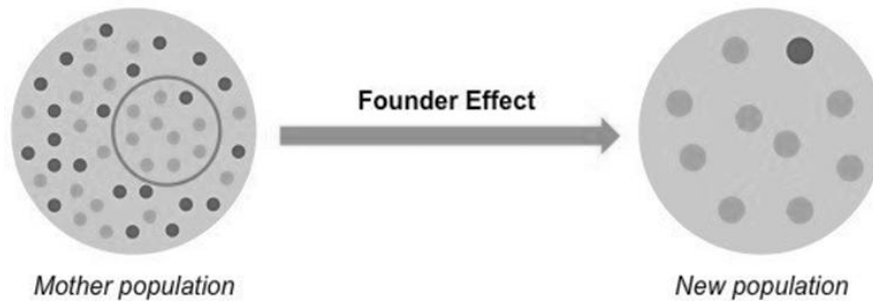


Figure 116. Founder effect

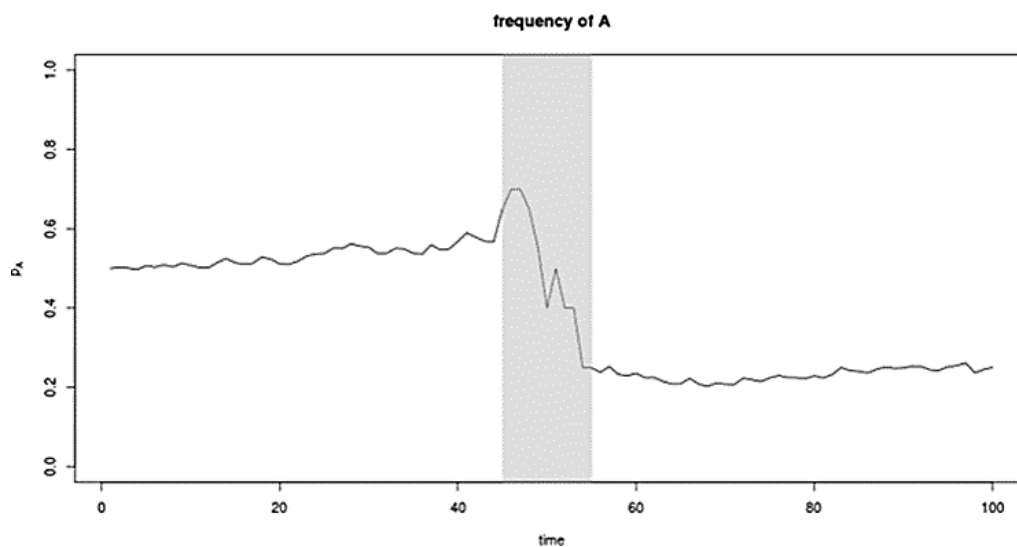


Figure 117. Simulation of the bottleneck effect. The x-axis is generations, and the y-axis is p (the frequency of the dominant allele). Alleles A and a are neutral and do not affect fitness. The initial population consists of 1000 individuals, and the frequency of allele A is 0.5. The number drops to 10 individuals for 10 generations and then recovers, with the frequency of allele A in the new population being about 0.2 due to genetic drift. A similar effect is observed when a new population is established by a small group of individuals (founder effect).

The resource <https://phytools.shinyapps.io/founder-event/> was used for modeling

Describing the random fluctuations of genotypes in the population and the loss of some of them, we can also consider the inheritance of the Y chromosome in males and mitochondrial DNA in the female line. Since not all fathers leave sons, and fathers who leave sons may have unequal numbers of them due to various random reasons, the frequency of some variants of Y chromosomes (haplotypes) changes with generations: some of them are lost as it happens with one of the alleles in Figure 2D, the frequency of others increases (Fig. 118).

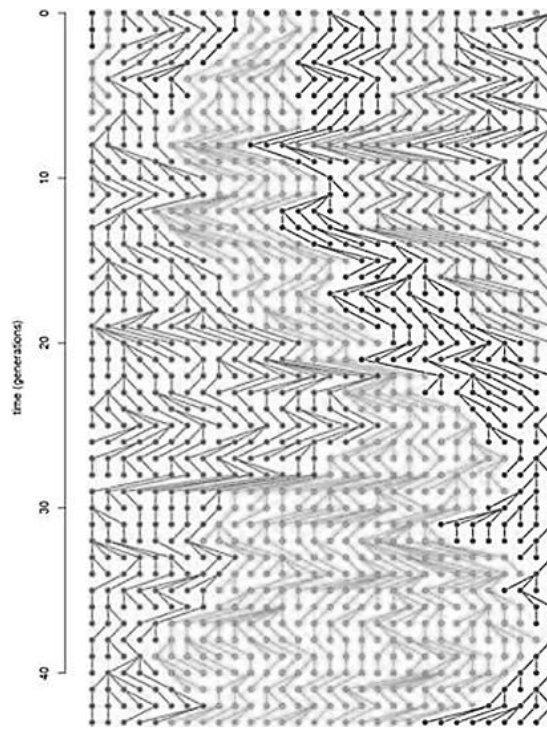


Figure 118. Principle of inheritance of mitochondrial DNA (or Y-chromosome). Since not all individuals leave daughters (sons), out of 30 variants in 1 generation only 3 remain by the 40th generation. Resource <https://phytools.shinyapps.io/coalescent-plot/> was used for modeling

If we analyze a large number of generations, the frequency of one of the haplotypes will inevitably reach 100 %. Indeed, it has been established that all modern human Y chromosomes come from one ancestral form that existed, according to molecular clock estimates, from 237 to 581 thousand years ago. A similar pattern is observed for mitochondrial DNA. Due to random fluctuations in frequencies, only one variant of it stays in the population. It has been established that the entire mtDNA variety of modern humans derives from the variant that existed about 234–99 thousand years ago (“lucky mother”). It should be noted that the above variants did not belong to the only humans on earth, as it is often wrongly interpreted, but were merely fixed in the population during random fluctuations in frequencies. Therefore, these data cannot be used as evidence for the existence of a single pair of the first humans on Earth.

Migration. In terms of population genetics, **migration** or **gene flow** is the transfer of genetic variants from one population to another which changes the composition of the gene pool of that population. **Immigration** is the inflow of individuals into a population, and **emigration** is the outflow of individuals from that population into another population. Migration reduces genetic differences between populations of a species.

Non-random mating. Another factor that changes the frequencies of genotypes in a population is non-random mating. Completely random mating (**panmixia**) is observed only in ideal populations. It is important to note that

non-random mating does not change allele frequencies, but it does change genotype frequencies.

Populations are called **panmictic** if there the choice of partner is not restricted. Most natural populations are non-panmictic because many factors prevent random mating.

Nonrandom mating has several forms:

- In **positive assortative mating**, similar genotypes are more likely to mate than dissimilar genotypes. This may occur in humans: a number of studies have shown that many people are more attracted to partners who physically resemble them and are therefore more likely to be genetically similar. A clinically important aspect of assortative interbreeding is the tendency to choose partners with similar medical problems, such as congenital deafness or blindness, or short stature.

An extreme manifestation of positive assortative crossing is self-pollination in plants. In this case, the frequency of heterozygotes will halve with each subsequent generation: dominant and recessive homozygotes in self-pollination produce progeny with the same genotype ($AA \times AA \rightarrow AA$, $aa \times aa \rightarrow aa$), but heterozygotes produce both heterozygotes and homozygotes ($AA \times Aa \rightarrow AA + Aa + Aa + aa$).

- **Negative assortative mating** occurs when dissimilar genotypes have a greater chance of mating. Examples are some plant species that have built-in recognition systems that prevent fertilization between individuals with the same alleles at certain loci.

- **Inbreeding** is the crossing between individuals who are related by descent. It increases the proportion of homozygotes and decreases the proportion of heterozygotes in the population. Theoretically, a fully inbred population will consist only of homozygous genotypes.

High levels of inbreeding may have negative effects because it increases the percentage of individuals homozygous for deleterious recessive mutations. The decrease in viability of individuals resulting from inbreeding is called **inbreeding depression**. An **inbreeding coefficient (F)** is used to describe the intensity of inbreeding in a population. This factor reflects the probability that two alleles of a given gene present in an individual are identical because they derive from an allele from the same ancestor.

CHARACTERISTICS OF SMALL HUMAN POPULATIONS

Human populations of more than 4,000 people are considered large. There are small populations of 1,500 to 4,000 people. Intra-group marriage in them is 80–90 %, and the inflow of new genes from other groups is 1–2 %. The smallest populations of people are the groups having up to 1.5 thousand people. Intragroup marriages account for over 90 %, and the influx of new genes from other groups is less than 1 %.

Due to the small number of individuals, in small populations the effect of genetic drift becomes strong. Because of this, some alleles have a greater chance to reach 100 % frequency. It may let the mutations that reduce fitness reach 100 % frequency. This can lead to the accumulation of mutations with negative effects. In addition, a small population size also limits the choice of a partner for marriage, which increases the level of inbreeding in the population and may also lead to a decrease in genetic diversity. Thus, a decrease in population size has a negative effect.

Genetic load. Genetic load is the difference between the fitness of the average genotype in a population and the fitness of some reference genotype which may be either the best available in the population or the theoretically optimal genotype. An average individual taken from a population with a low genetic load will usually have more surviving offspring when developed under the same conditions than an average individual from a population with a high genetic load. A high genetic load can put a population at risk of extinction. The most important contribution to the genetic load is the mutation load — the part of the genetic load resulting from the accumulation of deleterious mutations.

CHAPTER 14

HUMAN GENETICS AND MEDICAL GENETICS

A HUMAN BEING AS A SPECIFIC OBJECT OF GENETIC ANALYSIS

Human genetics studies the laws of inheritance of normal and pathological human traits and the dependence of their manifestation on the genotype and environmental factors.

The branch of human genetics studying the role of hereditary factors in human pathology is called medical genetics.

The relevance of medical genetics is extremely high. As of 2021, more than 6,000 hereditary diseases are known (according to The Genetic and Rare Diseases Information Center database). Genetic diseases and birth defects occur in about 2–5 % of all live births, are associated with up to 30 % of hospitalizations in pediatrics, and cause about 50 % of infant deaths in industrialized countries. Of recognized pregnancies, 15 % result in spontaneous abortions, up to 60 % of which are due to chromosomal abnormalities; among children born alive, 1 in 200 have chromosomal abnormalities (G. H. Valentine, 1979).

Medical genetics studies the role of heredity in human pathology, patterns of inheritance of hereditary diseases, as well as methods of their prevention, diagnosis, and treatment.

The tasks of medical genetics are:

- the study of pathogenesis, clinic, diagnosis, pharmacological and other treatments (gene therapy), and prevention of human hereditary diseases;
- early diagnosis of hereditary diseases by improving express methods and prenatal diagnosis;
- studying the mechanisms of hereditary predisposition and innate resistance to multifactorial diseases;
- studying genetic aspects of immunity, allergies, transplants, carcinogenesis, genetic engineering, etc.;
- the widespread introduction of medical genetic counseling into medical practice.

The study of human genetics is associated with some peculiarities and objective difficulties:

- complex genome — more than 20,000 protein-coding genes, 46 chromosomes, and 24 linkage groups;
- a slow change of generations — a period of child-bearing in humans is 18–35 years, a doctor-geneticist may observe not more than 3 generations during his life;
- the small number of descendants — most families have 1–2 children;
- impossibility to apply hybridological method; a doctor has no right to interfere in the formation of married couples, but if the doctor is interested in

young people already married, the doctor can examine them and prevent the inheritance of pathological traits in descendants;

– the impossibility of creating the same living conditions.

Despite these difficulties, human genetics is now better understood than the genetics of many other organisms. This has been facilitated by the social nature of humans, the growing needs for medicine, and the variety of modern research methods.

To examine patients and solve questions of hereditary and congenital disease pathogenesis, medical genetics widely uses common methods of clinical examination: ultrasonography, electrocardiography, electroencephalography, electromyography, biochemical analyses of biological fluids, tissue biopsy, etc. However, some specific methods can be used to study the emergence, development, spread, and mechanisms of intergenerational transmission of hereditary diseases and the role of genotype and environmental factors in their manifestation.

Classification of human genetics methods:

- Basic methods: genealogical method; twin study; cytogenetic methods; biochemical tests; population genetics.
- Rapid Diagnostic Tests (RDT): sex chromatin determination; biochemical tests; dermatoglyphic analysis; bacterial inhibition assays.
- Methods of prenatal diagnosis: alpha-fetoprotein test; ultrasonography; chorionic villus sampling (CVS), amniocentesis, cordocentesis, fetoscopy.

THE GENEALOGICAL ANALYSIS

The genealogical method was proposed in 1883 by F. Galton. It is based on the drawing of pedigrees (family trees) and tracing in a series of generations the transmission of a certain trait. The method allows determining:

- whether the trait is hereditary (according to its manifestation in the members of the pedigree); a pattern of inheritance (dominant or recessive, autosomal, X-linked or Y-linked);
- zygosity of family members (homo- or heterozygotes);
- penetrance of the gene (frequency of its expression);
- probability of giving birth to a child with a hereditary pathology (genetic risk).

Stages of genealogical analysis:

1. Collecting data from the propositus about all his or her relatives (anamnesis).
2. Drawing of pedigree (Fig. 119).
3. Analysis of the pedigree and conclusions.

The difficulty of collecting the anamnesis is that the propositus must know well the maternal and paternal relatives of at least three generations and their state of health.

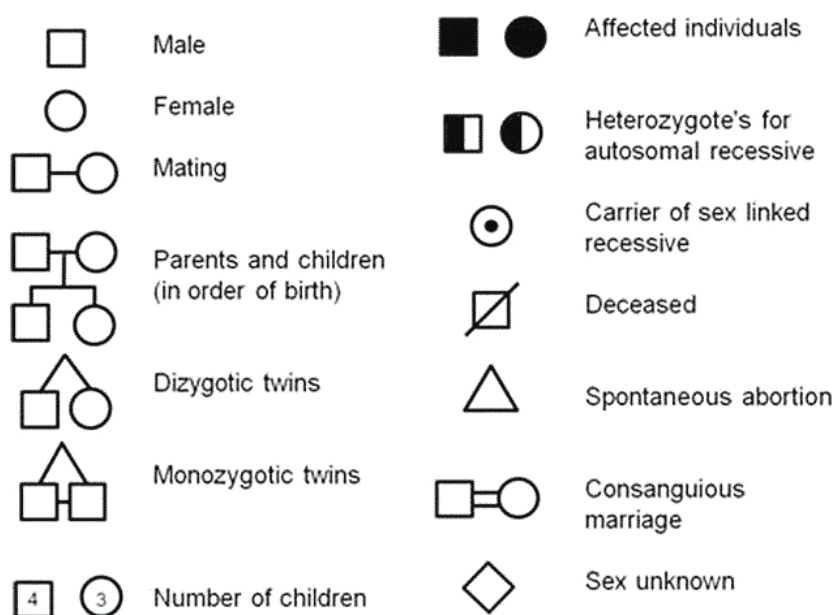


Figure 119. The symbols used in the construction of pedigrees

For the drawing of pedigrees, the legend proposed in 1932 by G. Just is used.

The basis of the pedigree is the *propositus*, the person with whom the family examination begins. In pedigrees, the *propositus* is marked with a sign $\nearrow$.

The genealogical method of research makes it possible to establish the patterns of inheritance of traits in a person.

Autosomal dominant inheritance. Autosomal dominant inheritance is characterized by the following features:

- the trait appears in every generation; a sick child is born by sick parents; men and women are equally affected (Fig. 120);

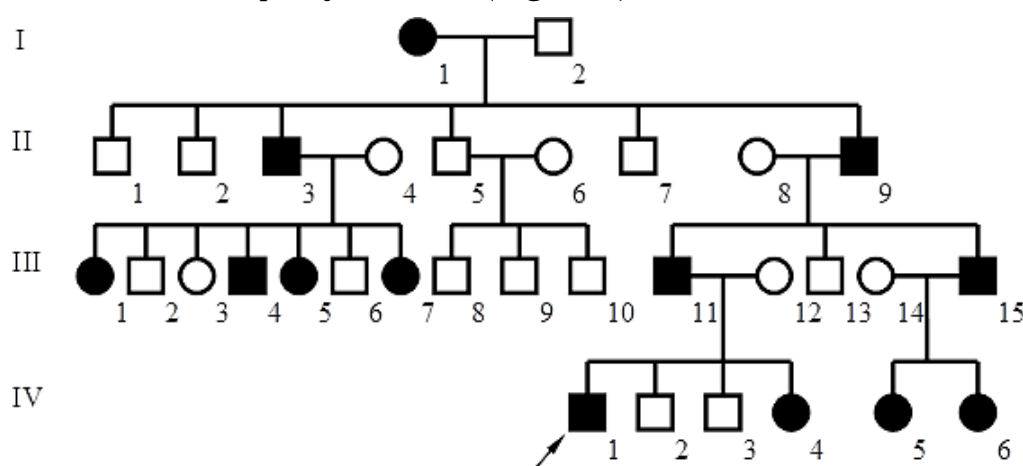


Figure 120. Pedigree with autosomal dominant type of inheritance

- trait inheritance is vertical and horizontal (manifestation of the trait is observed in the vertical and horizontal direction from the trait holder within the pedigree);

- 100 % probability of inheritance (if one parent is homozygous), 75 % (if both parents are heterozygous), and 50 % (if one parent is heterozygous and the other parent is recessive homozygote).

It should be emphasized that the above traits of autosomal dominant inheritance will only appear if complete dominance and 100 % gene penetrance are observed. Thus, polydactyly, freckles, curly hair, brown eye color, etc. are inherited in humans. In the case of incomplete dominance, the descendants will show an intermediate form of inheritance. If the penetrance is incomplete, there may not be patients in every generation.

Autosomal recessive inheritance. Autosomal recessive inheritance is characterized by the following features:

- the trait does not appear in every generation (Fig. 121);

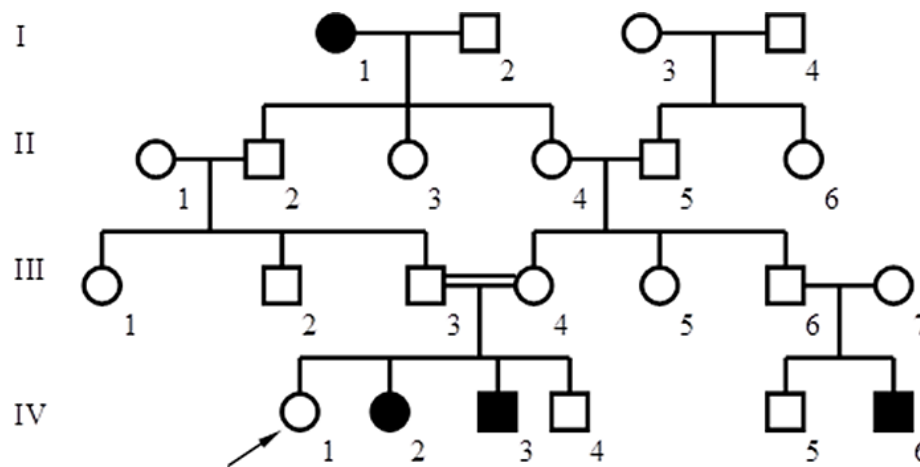


Figure 121. Pedigree with autosomal recessive type of inheritance

- a sick child (homozygote) is born by healthy parents (heterozygotes); males and females are equally affected;
- trait inheritance is mostly horizontal;
- 25 % probability of inheritance (if both parents are heterozygous), 50 % (if one parent is heterozygous and the other is recessive homozygous), and 100 % (if both parents are recessive homozygotes). Most often, the probability of inheriting autosomal recessive diseases is 25 %, because due to the severity of the disease, such patients (recessive homozygotes) either do not survive to childbearing age or do not marry.

This is how humans inherit phenylketonuria, sickle cell anemia, blue eyes (from the HERC2 gene), etc.

X-linked recessive inheritance. X-linked recessive inheritance is characterized by the following features:

- the trait does not appear in every generation; men are mostly affected;
- a sick child is born by healthy parents (Fig. 122);
- trait inheritance is predominantly horizontal;

– the inheritance probability is 25 % of all children, including 50 % of boys (if the mother is heterozygous and the father is healthy, which is most often the case).

This is how hemophilia, color blindness, hereditary anemia, muscular dystrophy, etc. are inherited in humans.

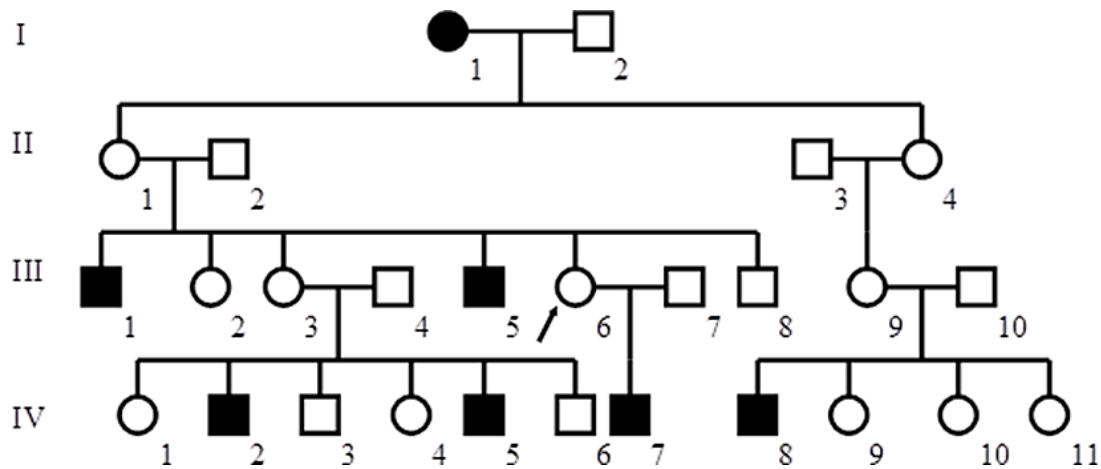


Figure 122. Pedigree with X-linked recessive type of inheritance

X-linked dominant inheritance. X-linked dominant inheritance is similar to the autosomal dominant type, except that the male transmits the trait to all his daughters but none of his sons (the sons receive the Y chromosome from the father, Fig. 123). An example of such a disease is a particular form of rickets that is resistant to vitamin D treatment.

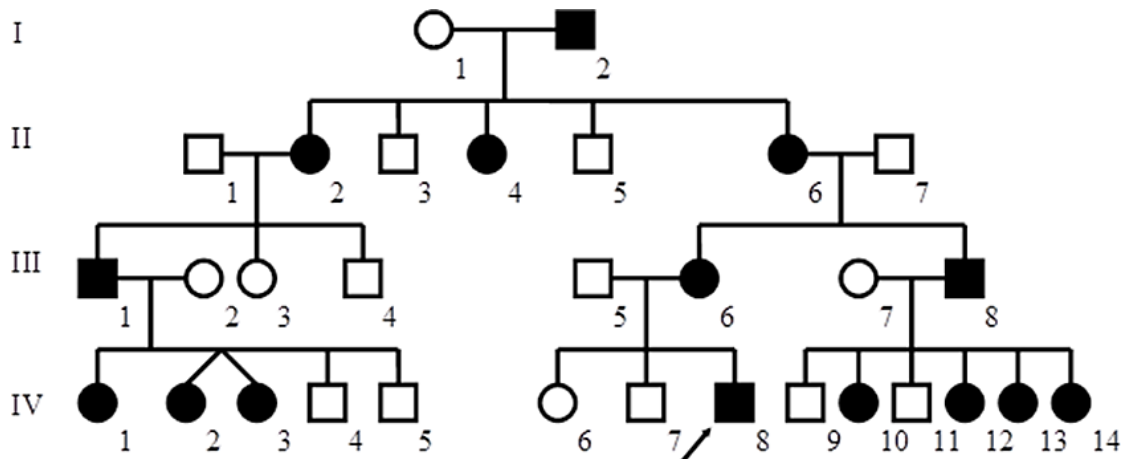


Figure 123. Pedigree with X-linked dominant type of inheritance

Y-linked inheritance. Holandric (Y-linked) inheritance is characterized by the following features:

- the trait appears in all generations;
- only men are affected;
- the affected father transmits the disease to all his sons (Fig. 124);
- the probability of inheritance in boys is 100 %.

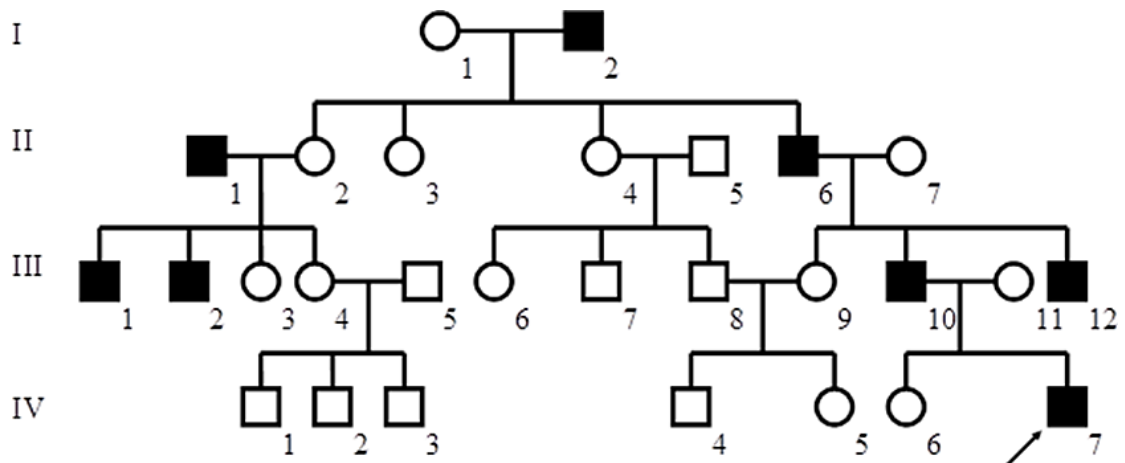


Figure 124. Pedigree with Y-linked type of inheritance

Some forms of ichthyosis, hypertrichosis (excessive hair growth) of the external auditory canals and middle phalanges, some forms of syndactyly (toe fusion), etc. are inherited in humans.

When analyzing a pedigree, the pattern of inheritance is determined by excluding other types of inheritance. For example, if a trait does not occur in every generation, the dominant patterns of inheritance are excluded.

TWIN STUDY

Twin study allows quantifying the degree of influence of heredity and environment on the formation of the studied trait. It is based on the similarity and differences between monozygotic and dizygotic twins.

A distinction is made between mono- and dizygotic twins. Monozygotic (MZ) or identical twins develop from the same ovum fertilized by a single sperm. They have virtually the same genotype (not counting the small number of individual mutations that arise during embryonic development) but may differ in phenotype, which is considered to be due to the influence of environmental factors.

Dizygotic (DZ) or fraternal twins (also referred to as “non-identical twins”) develop from two different zygotes formed after fertilization of simultaneously matured ova by different spermatozoa. Such twins have different genotypes and their phenotypic differences are due to both genotype and environmental factors. Monozygotic twins have a greater degree of similarity in traits that are determined mainly by genotype. For example, they have the same sex, blood types according to different systems (AB0, Rh, MN, etc.), eye color, the same dermatoglyphic patterns on their fingers, palms, and feet, etc. These signs are used as criteria for the determination of twin zygosity. For example, the difference between twins according to at least one of the criteria may mean that they are dizygotic.

The similarity of twins according to the studied trait is called concordance, and the differences are called discordance. Since monozygotic twins have the same genotype, their concordance in various traits is higher than that of dizygotic twins. For example, if schizophrenia develops or does not develop in both twins in 70 % of the pairs, then their concordance is 70 %. If the same is true for dizygotic twins in 13 % of cases, then their concordance is 13 %.

Stages of twin study:

1. Drawing a sample of twins.
2. Determination of twin zygosity.
3. Determination of the concordance degree of twin samples on the studied trait.
4. Calculation of the coefficient of inheritance.

To estimate the role of heredity and environment in the development of a particular trait Holzinger formula is used:

$$H = \frac{\text{CMT \%} - \text{CDT \%}}{100 \% - \text{CDT \%}}$$

where H is the degree of genetic determination, CMT is the % of concordant monozygotic twins in the group under study, CDT is the % of concordant dizygotic twins.

The degree of environmental influence (E) can be calculated by the formula:

$$E = 1 - H.$$

If the result of calculations according to Holzinger's formula is close to 1, the main role in the development of a trait belongs to heredity. The closer the result is to zero, the greater the role of environmental factors.

It should be emphasized that human morphological traits are more susceptible to genetic control than mental features (Fig. 125). This is evidenced by the analysis of an intelligence quotient (IQ) in monozygotic twins. When raised together, its concordance is 91 %, and in twins who grew up in different conditions — 67 % (in dizygotic — 64 %).

Disease	Monozygotic	Dizygotic
Celiac disease	75–83%	11%
Psoriasis	67%	15%
Primary biliary cirrhosis	60%	Not available
Ankylosing spondylitis	50%	20%
Systemic lupus erythematosus	33%	2%
Crohn's disease	25%	7%
Multiple sclerosis	25–31%	3–5%
Type 1 Diabetes	21–70%	0–13%
Ulcerative colitis	18.7%	3%
Graves' disease	17–31%	1.9–4.7%
Rheumatoid arthritis	12–15%	3.5%

Figure 125. Concordance of some traits in monozygotic and dizygotic twins

CYTOGENETIC METHODS

Cytogenetics is a branch of genetics that studies chromosomes. The most common method in human cytogenetics is light microscopy. Modern methods of karyotype analysis, such as comparative genomic hybridization and SNP karyotyping, provide karyotype data using DNA microarray technology.

Karyotyping. Karyotyping (cytogenetic examination) is the study of a patient's set of specifically stained metaphase chromosomes using a microscope to analyze their number and structure (Fig. 126).

Any tissue whose cells have nuclei and are capable of dividing can be used to study human chromosomes.

Preparation of chromosomes for microscopy using a venous blood sample as an example is as follows: a few milliliters of blood are added to a small volume of growth medium containing phytohemagglutinin. Phytohemagglutinin stimulates T-lymphocytes to divide. The cells are cultivated under sterile conditions at 37 °C for about 3 days, during which time they divide. Colchicine is then added to the culture. It stops mitosis at the metaphase stage by destroying the division spindle. A hypotonic solution is next added, which causes cell lysis and releases the chromosomes, which are then fixed, placed on a slide, and stained.

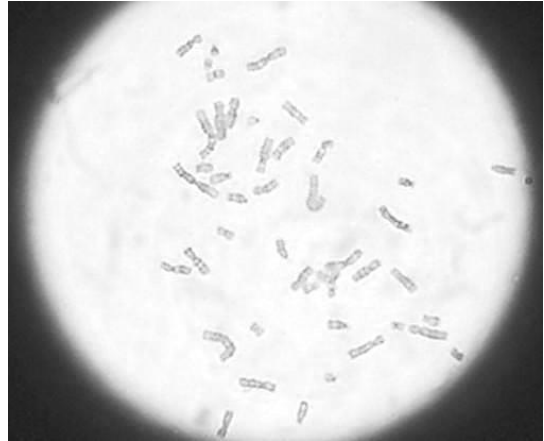


Figure 126. Giemsa-stained chromosomes under a light microscope

There are different methods of chromosome staining.

Simple staining, which was used until the early 1970s, allows the determination of only the chromosome group (see Denver classification of human chromosomes, Chapter 3), but not exact number.

Differential staining reveals on each human chromosome a typical pattern of dark and light bands, so it is possible to recognize each chromosome accurately (see Paris Nomenclature of Human Chromosomes, Chapter 3).

Fluorescent in situ hybridization. Fluorescent in situ hybridization (FISH) allows to determine the localization of the desired DNA sequence in chromosomes using a DNA probe. The DNA probe is visualized by fluorescent dye. The probe hybridizes only to a complementary DNA sequence at a specific location on one of the chromosomes. The patient's chromosomes can be visualized under a fluorescence microscope.

FISH is important for the clinical diagnosis of various chromosomal abnormalities, including deletions, duplications, and translocations.

Since a somatic cell contains two homologous chromosomes, in a healthy human the probe hybridizes in two places and two fluorescent “spots” are

observed under the microscope. This allows FISH to detect the loss of a chromosome segment: if the probe hybridizes with only one of the patient's chromosomes, there is likely deletion in the chromosome with which the probe does not hybridize. Therefore, FISH is often used to detect common syndromes often caused by deletions, such as Prader-Willi syndrome (microdeletion 15q11.2) and Williams syndrome (microdeletion 7q11.2). Additional copies of a chromosomal region can also be detected using FISH. In this case, the probe is hybridized in three or more locations instead of two.

Spectral karyotyping. Spectral karyotyping (SKY) is a technique based on hybridization with probes labeled with different fluorochromes, which allows the visualization of all chromosome pairs in different colors simultaneously. As a result, translocations can be easily detected because sections of different chromosomes are presented and stained in different colors.

To perform SKY, a set of fluorescently labeled probes is made for each chromosome.

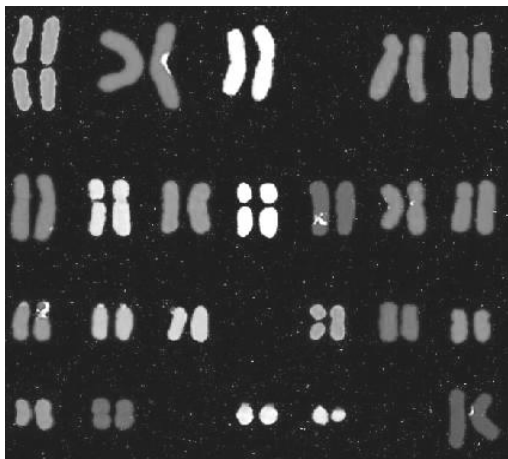


Figure 127. Visualization of chromosomes using SKY. Since each chromosome has a different color, any rearrangements are easily visualized

Metaphase chromosome samples are hybridized with probes and then stained. Each chromosome acquires unique spectral characteristics. To distinguish between them, an interferometer is used to read the image, and the image is analyzed using special software. It classifies chromosomes according to the obtained spectral characteristics and automatically sorts them according to chromosome nomenclature. In addition, by displaying an image of the chromosomes, the software assigns a different color to each of them, which facilitates their visual analysis (Fig. 127).

The advantages of SKY are the ability to detect multiple chromosomal rearrangements.

SNP-A KARYOTYPING

Single-nucleotide polymorphism array (SNP-A) karyotyping is a technology that allows evaluating the entire karyotype by using microarrays containing a large number of probes specific for different SNPs (single-nucleotide polymorphisms) contained in all chromosomes. Then, using special software, the entire genome is evaluated (it is possible to identify the addition of genetic material of a certain chromosome or its loss, uniparental disomy, etc.).

Biochemical methods. Biochemical methods make it possible to diagnose inherited metabolic disorders. Usually, such disorders are caused by mutations in certain genes, resulting in a decrease or loss of function of certain enzymes or transport proteins involved in the metabolism of certain substances. Therefore, the presence of a particular monogenic disorder can be detected by a decrease or increase in the concentration of certain metabolites in the urine, sweat, plasma, serum, blood cells, etc. or by a decrease in the function of certain enzymes. Biochemical methods can be qualitative or quantitative. Qualitative methods indicate the presence or absence of a substance in a sample, while quantitative methods measure its content in the sample.

An example of a qualitative method is the urine ferric chloride test to detect phenylketonuria: the appearance of the olive-green coloring of urine after 2–3 minutes after adding a few drops of 5–10 % solution of FeCl_3 and acetic acid to it. Quantitative methods used to confirm the diagnosis of this disease are based on determining the amount of phenylalanine in blood serum or the activity of phenylalanine-4-hydroxylase. Qualitative methods are more often used at the first stages of examination, when hereditary pathology is suspected, and are also more often used as screening methods (will be discussed later).

More complex and highly accurate quantitative methods play a decisive role in the diagnosis of metabolic disorders (chromatographic methods, tandem mass spectrometry).

Modeling. Biological modeling is the practice of studying certain properties of living organisms (e.g., hereditary diseases) on other biological objects. These can include animals, organoids (not to be confused with organelles!), and cell cultures. For example, hemophilia is found in dogs; in mice, cleft lip and cleft palate, similar to a similar anomaly in humans; in hamsters and rats, diabetes mellitus, achondroplasia, muscular dystrophy, etc. Although mutant lines of animals do not give an exact picture of human hereditary diseases, even partial reconstruction of their fragments in some cases allows the mechanisms of primary deviation from the norm to be revealed and methods for diagnosis, prevention, and treatment of these diseases to be developed in experiments.

With the appearance of human organoids, systems of three-dimensional cell cultures derived from stem cells, and functioning as human organs, it has become possible to recreate the architecture and physiology of human organs in great detail. Human organoids provide unique opportunities to study human diseases and complement animal models. Human organoids have been used to study infectious diseases, genetic disorders, and cancer. Moreover, recent studies have revealed biological processes that are specific to the human body and cannot be modeled on other animals. These include, for example, brain development, metabolism, and drug efficacy testing. Therefore, the emergence of in vitro 3D human cell culture approaches using stem cells from various organs has attracted widespread attention as being able to overcome these limitations.

Mathematical modeling involves creating mathematical models to explain biological processes. It is inseparably connected with computer modeling, without which it is impossible to work with the huge amount of data that modern biology deals with. Computer modeling can be used to analyze the complex network of biochemical reactions and transport processes of the cell, the transmission of signals in the cell, and the interaction of various molecules. An important task of such modeling is the prediction of the spatial structure of a protein from its amino acid sequence.

Computer models are used to develop new medicines and to predict their pharmacokinetics and pharmacodynamics.

RAPID DIAGNOSTIC TESTS (RDT)

Rapid diagnostic methods are quick preliminary methods for diagnosing human inherited diseases. They are often used to screen large populations of people for hereditary pathologies as screening methods (screening programs). For example, screening newborns for phenylketonuria and hypothyroidism; pregnant women — for the content of α -fetoprotein in the blood, which suggests some fetal malformations.

There are certain requirements for express methods:

- the method must be diagnostically significant, i.e., give false-positive and false-negative results as rarely as possible;
- the method must be reliable: the same sample must give the same result when independently tested twice;
- small quantities of readily available material (blood, urine) should be tested (e.g., capillary blood stains dried on filter paper); the method must be acceptable to the subjects and medical staff;
- the method must be economical.

Bacterial inhibition assays. Bacterial inhibition assays (microbiological tests) are used to detect in a biological sample certain substances necessary for the growth of certain strains of microorganisms (amino acids, sugars, etc.). An example of such tests is the Guthrie bacterial inhibition assay used to diagnose phenylketonuria, histidinemia, galactosemia, and maple syrup urine disease. The Guthrie test has been used as a newborn screening test. Let us consider the principle of this method using phenylketonuria as an example. Patients with this disease have significantly elevated concentrations of phenylalanine and its derivatives in the blood, which is what the test determines. A small amount of blood is taken from the heel of the newborn and applied to a special sheet of filter paper. Next, the paper disc soaked in the blood sample was placed on nutrient media. The nutrient media contains *Bacillus subtilis* bacteria as well as β -2-thienylalanine, a substance that inhibits the growth of these bacteria, an antimetabolite of phenylalanine. However, this

effect of β -2-thienylalanine disappears in the presence of a sufficiently high concentration of phenylalanine in the medium. Thus, if the tested blood sample contains phenylalanine, bacterial growth begins around it, which is regarded as a positive test result (Fig. 128).

Similarly, the presence of other amino acids and carbohydrates in the blood can be diagnosed using various antimetabolites and appropriate bacterial strains.

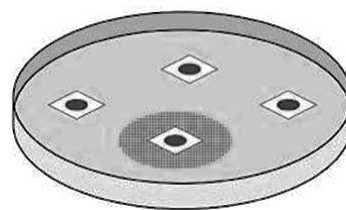


Figure 128. Guthrie bacterial inhibition assay used to diagnose phenylketonuria

Rapid biochemical tests. The biochemical tests determine metabolites specific to inherited metabolic disorders (enzymopathies). Such methods are often used for population screening. Population screening is large-scale testing of a population to identify individuals who are likely to have the disease and those who probably do not. Screening tests are not intended to make a definitive diagnosis but are intended to identify subpopulations on which further, more accurate diagnostic tests should be performed. Examples of rapid methods used for screening are tests such as the urine ferric chloride test for phenylketonuria, which determines phenylalanine products in the urine, the sweat test using analyzers that allow quick and accurate measurement of sweat chlorides to diagnose cystic fibrosis, etc. More advanced methods, such as chromatography, mass spectrometry, etc., are used to confirm diagnoses.

X- and Y-chromatin determination. Determination of X- and Y-chromatin are diagnostic tests that determine the genetic sex. Oral mucosa cells are most commonly used for this test. Smears are stained with aceto-orcein and studied using light microscopy. The female karyotype must contain two X chromosomes, one of which is inactivated and represented as a Barr body — a clump of heterochromatin attached to the nucleus membrane (absent in men). Due to this mechanism, only one X chromosome functions in both male and female cells. Thus, it is possible to determine the number of X chromosomes in the karyotype — always 1 more than the number of Barr bodies determined.

To determine male Y-sex chromatin, smears are stained with quinacrine and examined using luminescence microscopy. This method allows determining the number of Y chromosomes in the karyotype — it is equal to the number of luminous dots in the nucleus. Determination of X- and Y-chromatin is a screening method; the final diagnosis of the chromosomal disorders is made after the karyotype is examined.

Dermatoglyphic analysis. The dermatoglyphic analysis is the study of the papillary patterns of the fingers, palms, and feet. Dermatoglyphic features often accompany certain chromosomal diseases and congenital anomalies are known. One of them is the single transverse palmar crease. Normally, there are three main flexural creases on the palm: the thumb crease, the oblique crease, and the transverse crease. The single transverse palmar crease is formed when

the oblique crease merges into one with the transverse crease. Its incidence in healthy people does not exceed 5 % but increases significantly in certain forms of pathology, such as Down syndrome. Other dermatoglyphic features characteristic of hereditary pathology are also known, but at present, due to the increasing availability of molecular genetic methods, the relevance of dermatoglyphic analysis is low.

Newborn screening for monogenic diseases. Screening is a mass examination of target populations aimed at detecting patients with certain diseases or carriers of mutant alleles. The aim of screening programs may be the prevention of disease, its early diagnosis and timely initiation of treatment, and prevention of serious complications of the diseases under study.

Newborn screening is a mass examination of newborns, the purpose of which is to identify children who have certain diseases at a sufficiently early stage (before the appearance of characteristic symptoms), which allows initiating therapy and preventing the development of severe consequences of the disease. Screening programs do not make a definitive diagnosis but only identify individuals who are presumed to have the disease. The diagnosis is confirmed by a more detailed examination.

Requirements for newborn screening programs for hereditary diseases:

- The disease must be well studied clinically and laboratory.
- The frequency of the disease in the population must be relatively high.
- The disease must have severe consequences, leading to disability and death, but there must be an effective treatment for the disease.
- Laboratory tests used for screening should not give false-negative results, and the ratio of true-positive to false-positive results should be at least 1 : 5;
- The tests used must be technically simple, safe, economical, and ethically acceptable.

The screening program consists of several steps:



Figure 129. Blood sampling of a newborn for screening for hereditary diseases

1. Obtaining newborn's blood samples (Fig. 129), transporting them to the laboratory, and performing the tests.
2. Conduct confirmatory laboratory diagnostics for positive results of the initial screening.
3. Initiation of early treatment of the detected disease to prevent the development of its severe consequences.
4. Genetic counseling of the family.

5. Screening programs differ from country to country, primarily in the number of diseases included in newborn screening. This is due both to the varying relevance of different diseases in different regions and to other characteristics of different countries.

CHAPTER 15

HEREDITARY DISORDERS

MONOGENIC METABOLIC DISORDERS

Any metabolic process is a chain of reactions catalyzed by enzymes encoded by genes. Mutations in such genes cause *inborn errors in metabolism*. The disorders caused by the mutation in a single gene are called *monogenic disorders*.

DISORDERS OF AMINO ACID METABOLISM

Classic **phenylketonuria** (PKU) is caused by mutations in the gene of *phenylalanine hydroxylase* (PAH). The disorder is inherited in an autosomal recessive pattern. The incidence frequency of PKU varies from ~ 1 in 10,000 people of European descent to 1 in 90 000 in those of African descent.

The enzyme phenylalanine hydroxylase converts phenylalanine to tyrosine. As a result of the enzyme defect, phenylalanine accumulates and is converted into other metabolites (e.g. phenylpyruvic acid, Fig. 130). Elevated levels of plasma phenylalanine have an adverse effect on essential cellular processes in the nervous system such as myelination and protein synthesis.

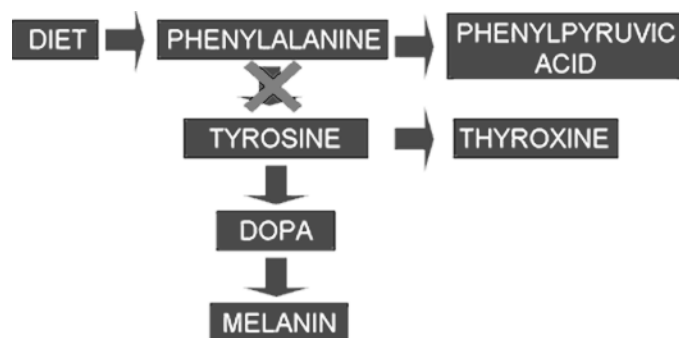


Figure 130. Metabolism of phenylalanine. Image from biocheminfo.wordpress.com

Clinical features: newborns with PKU initially don't have any symptoms. Without treatment, symptoms develop within a few months. The symptoms of PKU can be mild or severe and may include:

- Delayed development;
- Intellectual disability;
- A musty odor in the breath, skin, or urine;
- Neurological problems that may include seizures;
- Skin rashes (eczema);
- Fair skin, blond hair, and blue eyes.

Diagnosis is made based on the results of newborn screening (i.e. testing all newborns in their first days of life). Such screening is done by the Guthrie

test (see previous chapter) or other techniques analyzing molecules in a sample. Early diagnosis and treatment are essential to prevent developmental delays.

Treatment. The correction of the mutant gene is not currently possible; though intellectual impairment can be prevented by giving a low phenylalanine diet (Fig. 131). Phenylalanine is an essential amino acid and therefore cannot be removed entirely from the diet. The diet consists of a specialized medical formula in combination with regular foods that are low in phenylalanine. Such a diet is recommended lifelong.



Figure 131. A medical formula for patients with PKU and foods with high/low contents of phenylalanine

Oculocutaneous albinism (OCA) is caused by mutations in the gene encoding *tyrosinase* (or genes coding for several other enzymes involved in melanin production; in addition, there are other albinism types apart from OCA). OCA is inherited in an autosomal recessive manner. The worldwide frequency is 1 in 20,000 people.

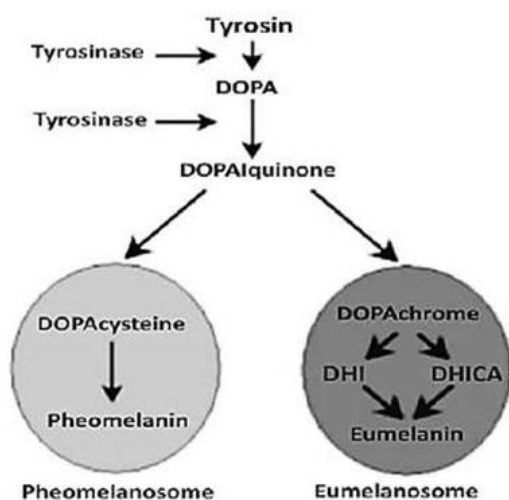


Figure 132. The production of melanin

Tyrosinase is necessary for the formation of melanin from tyrosine. OCA is characterized by complete or partial loss of melanin in the skin, hair, and eye (Fig. 132).

The lack of eye pigment results in poor visual acuity. Reduced fundal pigmentation causes underdevelopment of the fovea (a region of the retina) and abnormal routing of the nerve fibers from the eye to the brain.

Key diagnostic features are hypopigmentation of the skin and hair and characteristic ocular changes.

Treatment is based on the correction of symptoms:

- Correction of refractive errors with spectacles or contact lenses may improve visual acuity.
- Hats with brims and dark glasses or transition lenses often reduce discomfort in bright light.
- Protection from sun exposure with appropriate skin-covering clothing and sunscreens prevents burning, consequent skin damage, and the enhanced risk of skin cancer.
- It is recommended to avoid prolonged sun exposure.

DISORDERS OF CARBOHYDRATE METABOLISM

Galactosemia is the most common monogenic disorder of carbohydrate metabolism. Classic galactosemia caused by mutations in the gene encoding *galactose-1-phosphate uridyl transferase* (GALT) affects 1 in every 50,000 newborns. It is inherited in an autosomal recessive manner.

The enzyme activity is diminished so galactose is not effectively converted to glucose and, consequently, galactose is alternatively metabolized to galactitol and galactonate (Fig. 133). Galactose and its metabolites accumulate in many tissues. Other pathological physiology of classic galactosemia is not well understood.

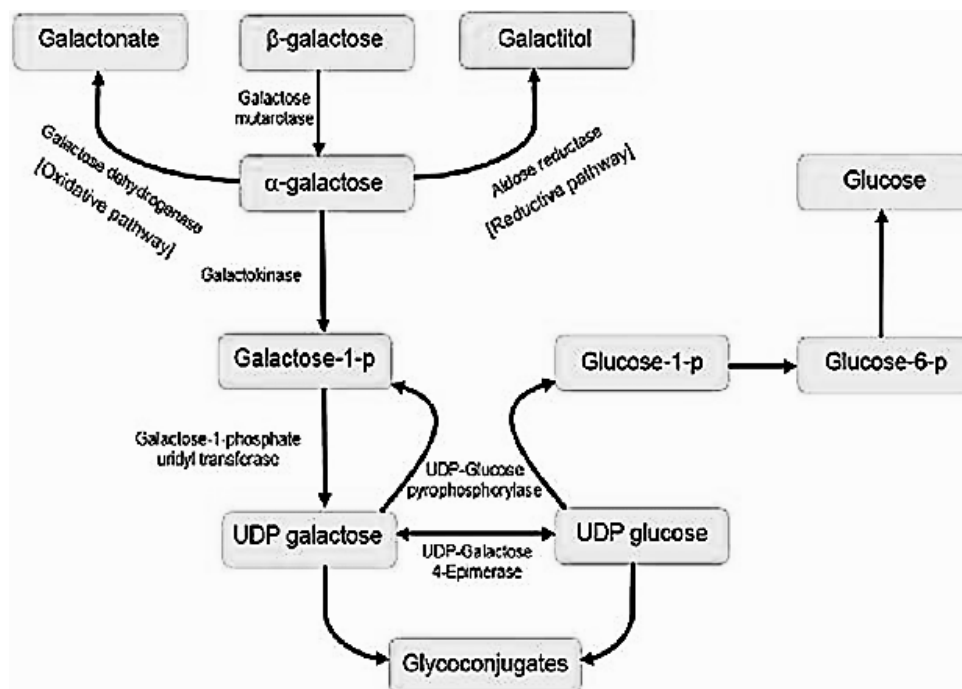
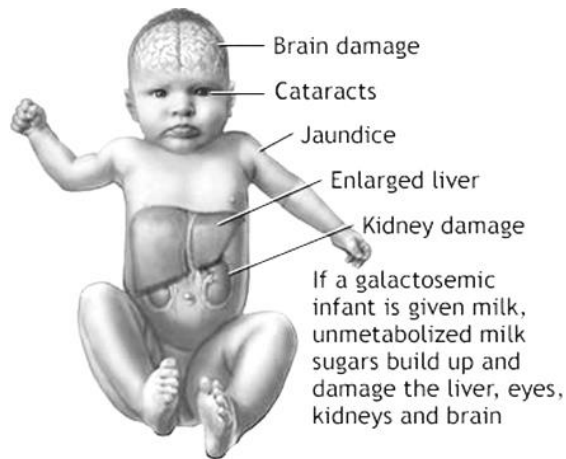


Figure 133. Biochemistry of galactose. Image from <https://www.wikidoc.org>

Clinical features (Fig. 134):

- Classic galactosemia typically manifests several days or weeks after birth.

- Symptoms include poor sucking, vomiting, failure to thrive (have problems eating and tend to lose weight), and jaundice.
- Cataracts (opacification of the lens of the eye) are found in about 10 % of infants.
- If galactosemia is left untreated, sepsis, hyperammonemia, and shock leading to death usually follow.



A

B

Figure 134. Galactosemia:
A — basic clinical features; B — cataract

Diagnosis can be confirmed by quantification of erythrocyte galactose-1-phosphate and galactose-1-phosphate uridyl transferase. Classical galactosemia shows < 1 % GALT activity and markedly increased Gal-1-P. Alternatively, diagnosis can be confirmed by genetic testing.

Treatment is the withdrawal of milk and, if symptomatic, emergency measures. People with galactosemia must avoid all milk products and other foods that contain galactose for life.

DISORDERS OF LIPID METABOLISM

Hyperlipoproteinemias are disorders of lipid metabolism characterized by abnormally elevated levels of lipids or lipoproteins in the blood. Hyperlipoproteinemias are of several types which can be acquired and congenital with autosomal dominant and autosomal recessive patterns of inheritance. Development of these disorders is often associated with mutations in genes coding proteins participating in lipid transport in the blood.

Hyperlipoproteinemias have a high clinical significance due to the increased risk of cardiovascular and cerebrovascular disease (e.g. heart attack, stroke).

Diagnosis is based on revealing abnormally elevated levels of lipids or lipoproteins in the blood.

Treatment depends on the type of hyperlipoproteinemia and commonly includes a low-fat diet and medicines to lower lipid levels in the blood.

DISORDERS OF PURINE METABOLISM

Lesch–Nyhan Syndrome (LNS) is caused by mutations in the gene coding for the enzyme *hypoxanthine-guanine phosphoribosyltransferase* (HPRT). The prevalence of Lesch–Nyhan syndrome is approximately 1 in 380,000 individuals and the pattern of inheritance of X-linked recessive.

Because of the absence or deficiency of the HPRT activity, the purines hypoxanthine and guanine are not built into nucleotides. Hence, purines are not recycled, but broken down, producing abnormally high levels of uric acid. The high uric acid levels cause the formation of sodium urate crystals which may accumulate in the joints and kidneys. For unknown reasons, a deficiency of HPRT1 is associated with low levels of dopamine in the brain (a chemical messenger involved in the control of physical movement and emotional behavior). However, it is unclear how a shortage of HPRT1 causes the neurological and behavioral problems characteristic of Lesch–Nyhan syndrome.

Clinical features:

- The first symptoms of Lesch–Nyhan syndrome may develop at six months of age.
- Due to excessive amounts of uric acid in the urine, its orange deposits can be found in the diapers (“orange sand”, Fig. 135, A).
- Later other symptoms of gout develop, such as the presence of kidney stones, gouty arthritis (accumulation of uric acid causes inflammation of the joints), subcutaneous tophi (deposits of monosodium urate crystals under the skin, Fig. 135, C).
- This is accompanied by distinctive neurological symptoms such as impairment of the muscular tonus, inattention, impulsivity, hyperactivity, and behavioral disturbances including self-injury (Fig. 135, B).
- Intellectual disability may also occur and is typically moderate.

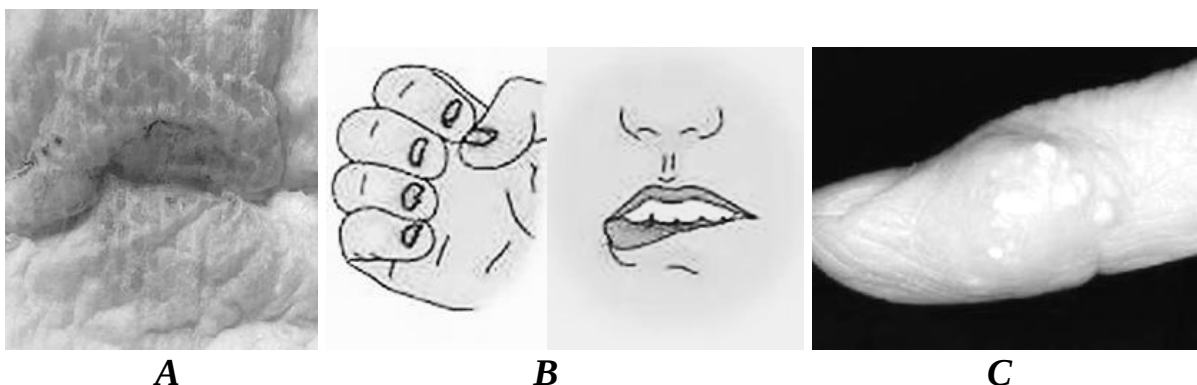


Figure 135. Clinical features of LNS:

A — urate crystals in a diaper; B — nail and lip biting — example of self-injury behavioral problem; C — tophus on the finger of a patient sick with gout

Diagnosis:

- Increased levels of uric acid in the blood and urine (but not always).
- Undetectable HPRT enzyme activity in peripheral blood or in intact cells (erythrocyte, fibroblast).
- The most reliable method is the identification of the *HPRT1* gene.

Treatment for LNS is symptomatic. Gout can be treated with allopurinol to control increased levels of uric acid. There is no efficient treatment for neurological dysfunction. The prognosis for individuals with LNS is poor. Death is usually due to renal failure in the first or second decade of life. With optimal care, few patients live beyond 40 years and most are confined to a wheelchair.

DISORDERS OF MINERAL METABOLISM

Wilson–Konovalov disease is an autosomal recessive metabolic disorder in which impaired copper excretion causes copper accumulation in the body. The prevalence of Wilson–Konovalov disease is 1 in 30 000. The disease is caused by mutations in the gene *ATP7B* coding for a protein that must normally incorporate copper into another protein called ceruloplasmin (Fig. 136). Impairment of this process causes low ceruloplasmin level in blood serum, reduced copper excretion with bile, and, in turn, accumulation of copper in tissues. In the initial stages, copper accumulates in the liver. As the disease progresses, copper deposits appear in other organs, most importantly in the brain and cornea.

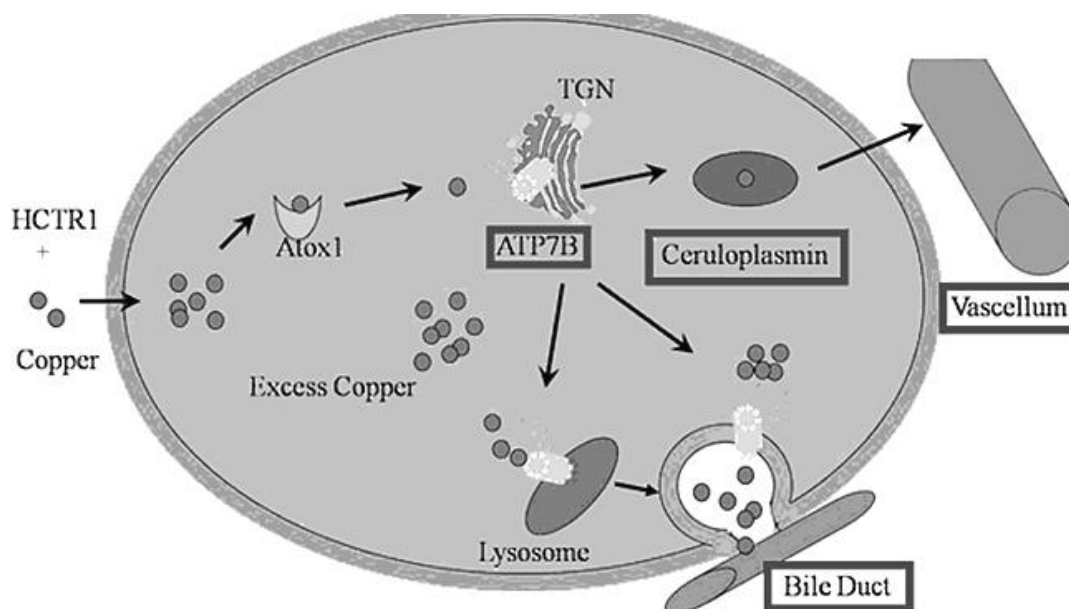


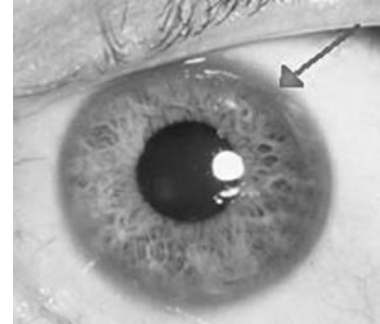
Figure 136. Diagram of ATP7B taking part in copper transport. Image from mdpi.com

Clinical features:

- The age of onset is from 5 to 35 years (more commonly 12–23 years).
- Damage to the liver may cause liver disease of different severity, including acute liver failure (a severe form of liver injury), acute or chronic hepatitis (inflammation of the liver), and cirrhosis (replacement of liver tissue with fibrous tissue). Some common symptoms are hepatomegaly, abdominal pain, and jaundice.

- Accumulation of copper in the brain causes neurological symptoms such as motor dysfunction of the tongue, lips, or vocal cords, slowness and diminution of amplitude or speed of movements, and involuntary shaking of limbs.

- Copper accumulation in the cornea results in 1–2 mm wide, brown rings in the periphery of the iris (Kayser–Fleischer rings, Fig. 137). The rings occur in ~ 98 % of patients who also have neurological symptoms, though their absence does not contradict the disease.



One of the best diagnostic signs of Wilson–Konovalov disease is the presence of Kayser–Fleischer rings. Though diagnosis is confirmed by a blood test which reveals the low concentration of ceruloplasmin accompanied by the high concentration of free copper in serum and increased daily copper excretion with urine.

Figure 137. Kayser–Fleischer ring.
Image from wikipedia.org

Treatment basically includes a low-copper diet and medicines that bind free metal ions.

DISORDERS OF BLOOD COAGULATION

Hemophilia A is an X-linked recessive caused by mutations in the gene coding for coagulation factor VIII (8th; gene F8). Hemophilia A is ~80 % of all hemophilia cases. There is also hemophilia B (~20 % of cases) caused by mutations in the gene coding for coagulation factor IX. Approximately 30 % of hemophilia A cases are caused by a new mutation without previous family history of the disorder. Hemophilia A — occurs in approximately 1 in 4,000–5,000 males.

Coagulation factors are the proteins needed for blood clotting (the process by which blood seals a wound to stop bleeding). The coagulation factors VIII and IX are made in the liver and circulate in the blood in inactive forms. Injury causes the activation of the coagulation factors and this leads to the initiation of a chain of chemical reactions that form a blood clot. Because of mutations abnormal forms of factor VIII are produced, or its amounts reduce. The altered

or missing protein cannot effectively participate in the process of blood coagulation and blood clots do not form properly.

Clinical features:

- The severity and symptoms of hemophilia can vary greatly from one person to another depending on the coagulation factor level (a mild, moderate, or severe form of the disease).
- Different degrees of trauma cause spontaneous or delayed onset bleeding (in joints (Fig. 138), muscular and soft tissue, and mucosa). Repeated bleedings in joints damage them.
- Mild hemophilia may be undiagnosed until a surgical procedure is needed or severe trauma occurs. Individuals with mild hemophilia may not experience their first bleeding episode until adulthood.
- In moderate hemophilia hematomas may occur after following mild trauma and in severe hemophilia hematomas may occur spontaneously. Individuals with the severe form of hemophilia can have spontaneous bleeding into any organ system including the kidneys (which may cause blood in the urine), the gastrointestinal tract (black or bloody stools), and the brain (which may cause neurological symptoms).
- Female carriers may show mild symptoms.



Source: Lichtman MA, Kipps TJ, Seligsohn U, Kaushansky K, Prchal JT: *Williams Hematology, 8th Edition*: <http://www.accessmedicine.com>
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Figure 138. Bleedings in joints in case of hemophilia

Diagnosis:

- Characteristic symptoms, a detailed patient history.
- Blood tests include complete blood count, coagulation tests, and measurement of the level of specific factors.
- Genetic testing.

The main treatment procedure is the substitution of coagulation factors. There is also gene therapy for hemophilia B.

DISORDERS CHARACTERIZED BY ABNORMAL HEMOGLOBIN STRUCTURE (HEMOGLOBINOPATHIES)

Sickle cell anemia is an autosomal recessive disorder caused by a point mutation in the gene *HBB*. This gene codes for the protein that is part of hemoglobin — β -globin (beta-globin). The mutation causes the replacement of one glutamic acid of β -globin with valine (Fig. 139) and, hence, produces an abnormal version of beta-globin known as hemoglobin S (HbS). HbS polymerizes when deoxygenated, causing deformation of erythrocytes (“sickling”, Fig. 140). This can be triggered by any event associated with reduced oxygen tension (hypoxia). In homozygotes, up to 100 % of the hemoglobin molecules are affected, leading to sickle cell formation under minimally decreased oxygen tension. In heterozygotes, sickling only occurs due to a severe reduction in oxygen tension. Sick cells break easier and have a reduced lifespan (120 $\rightarrow$ 10–20 days) which can lead to anemia (low count of hemoglobin and red blood cells). Sometimes the inflexible, sickle-shaped cells get stuck in small blood vessels and can cause complications.

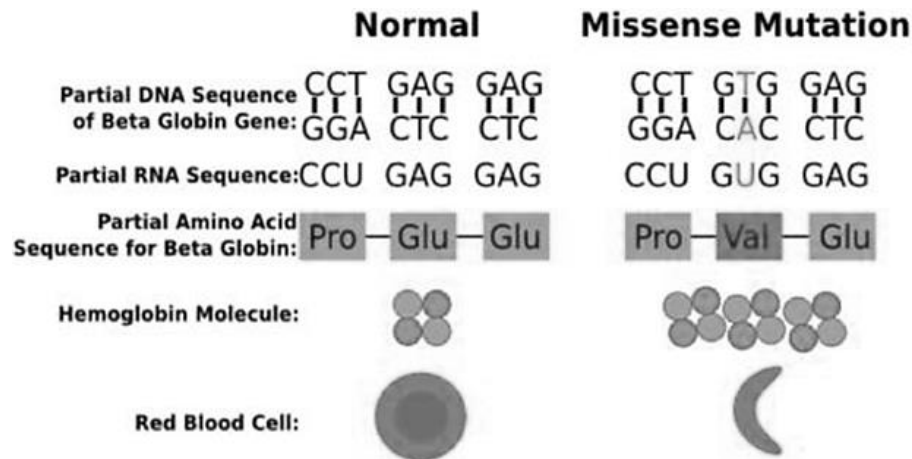


Figure 139. The point mutation causing HbS formation. Image from <http://tbl.med.yale.edu>

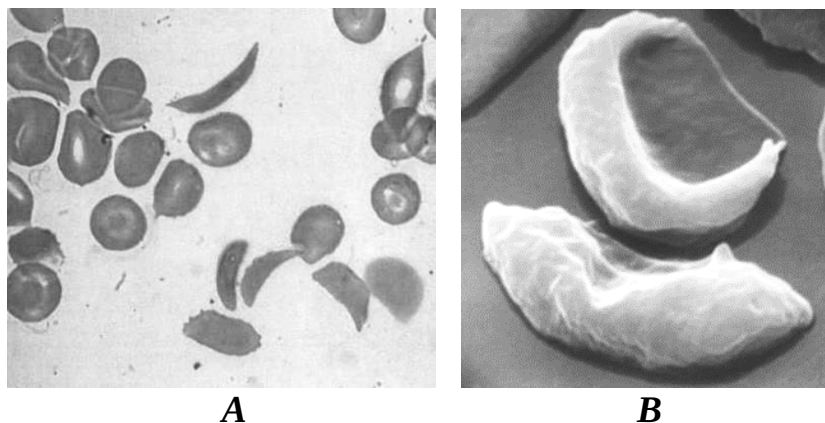


Figure 140. Sickle-shaped RBC:
A — light microscopy; B — electron microscopy

The disease is much more common in Africa (30 % heterozygote prevalence) as HbS provides resistance to malaria that is common in that geographic area.

Clinical features:

- Nearly 30 % develop symptoms in the first year of life; > 90 % by age 6 years.

- Anemia — fatigue, weakness, pallor.

- Block of blood flow (vaso-occlusion) in small blood vessels causes recurrent episodes of severe deep bone pain, acute chest syndrome, stroke, inflammation of the fingers and/or toes, and infarctions of virtually any organ with corresponding symptoms.

- Chronic pain.

- Frequent infections.

Diagnosis:

- Hemoglobin electrophoresis (gold standard).

- Sickle cells in a blood smear under anaerobic conditions.

The treatment is the prevention of vaso-occlusive crises (reduction of HbS production by medicines) and management of anemia, pain, and prevention of infections (vaccination, antibiotics).

CHROMOSOMAL ABERRATIONS

Chromosomal aberrations are abnormalities involving the structure (structural chromosomal aberrations) or the number of chromosomes (numerical chromosomal aberrations). They are the most common cause of spontaneous abortions (accounting for 60 % of cases).

Patau syndrome (Trisomy 13, the karyotype ♀: 47, XX,13+ or ♂: 47, XY,13+; Fig. 141) is caused by the presence of extra genetic material from the 13th chromosome in some or all cells of the body. The extra genetic material disrupts the normal course of development and causes the characteristic features of the syndrome. The incidence frequency is 1 in 16,000 newborns.

- Full trisomy 13 (the most common). Caused by the presence of 3 copies of the 13th chromosome in each somatic cell instead of 2 copies. This results from chromosomal nondisjunction during meiosis.

- Mosaic trisomy 13 is the less common case when only a percentage of cells contains the extra 13th chromosome whereas other cells contain the normal chromosomal pair (nondisjunction during mitosis).

- Partial trisomy 13 occurs when two full copies of chromosome 13 and an extra part of chromosome 13 are present. This can occur when part of chromosome 13 is translocated to another chromosome.

The full and mosaic trisomies are not inherited. The disorder can be inherited only from an unaffected parent who has balanced translocation (rearrangement of genetic material between chromosome 13 and another chromosome) as the extra genetic material may pass to children.

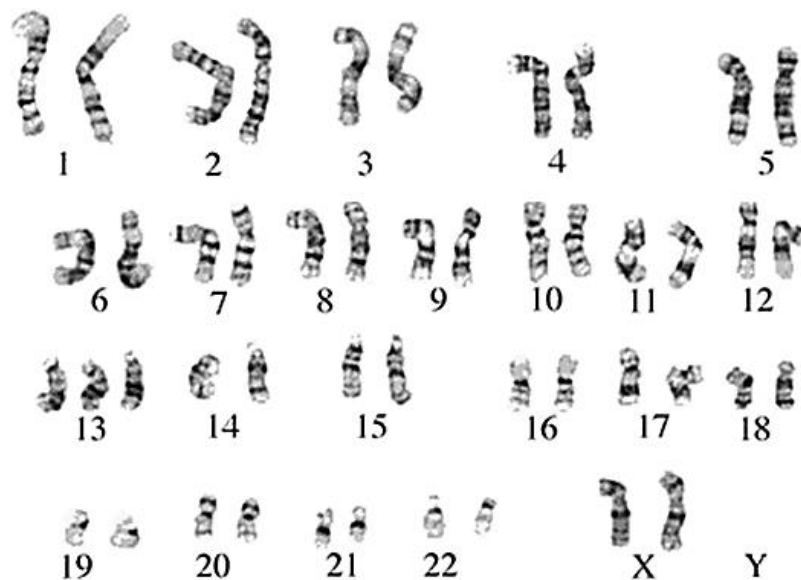


Figure 141. A female karyotype demonstrating trisomy 13

Clinical features (Fig. 142):

- Microcephaly (abnormally small head and underdeveloped brain), holoprosencephaly (failure of the anterior portion of an embryo's brain to develop into two cerebral hemispheres)
- Cleft lip and palate, low-set malformed ears, bulbous nose, small chin.
- Microphthalmia (abnormally small eyes), possibly coloboma (fissure in eye structures such as the iris).
- Polydactyly (the presence of more than 5 digits on a hand or foot), flexed fingers.
- Congenital heart defects (particularly ventricular septal defect, patent ductus arteriosus).
- Rocker-bottom feet (convex deformity of the plantar side of the foot).
- Visceral and genital anomalies, especially of the kidneys and ureter.
- Severe intellectual disability.

Diagnosis: commonly revealed during prenatal testing. Genetic testing can be done by cytogenetics tests (e.g. karyotyping) or molecular genetics tests.

There are no effective treatments. The only therapies used are directed toward the specific symptoms developing in each individual case. Sometimes this may include surgical correction of certain abnormalities caused by the disorder. Only 5 % of infants survive past 6 months of age.

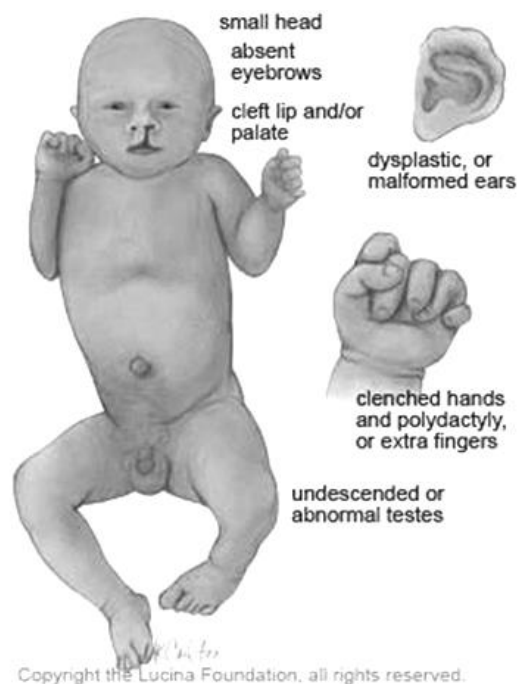


Figure 142. Some signs of Patau syndrome

Edwards syndrome (Trisomy 18, the karyotype ♀: 47, XX,18+; ♂: 47, XY,18+; Fig. 143) is caused by the presence of extra genetic material from the 18th chromosome in some or all cells of the body. The extra genetic material disrupts the normal course of development and causes the characteristic features of the syndrome. The incidence frequency is 1 in 5,000 newborns.

- Full trisomy 18 (3 copies of the 18th chromosome in each somatic cell instead of 2 copies because of chromosomal nondisjunction in meiosis).
- Mosaic trisomy 18 (only a percentage of cells contains the extra 18th chromosome because of nondisjunction during mitosis) occurs in ~5 % of cases.
- Partial trisomy 18 occurs in rare cases when the q-arm of chromosome 18 is translocated to another chromosome.



Figure 143. A female karyotype demonstrating trisomy 18

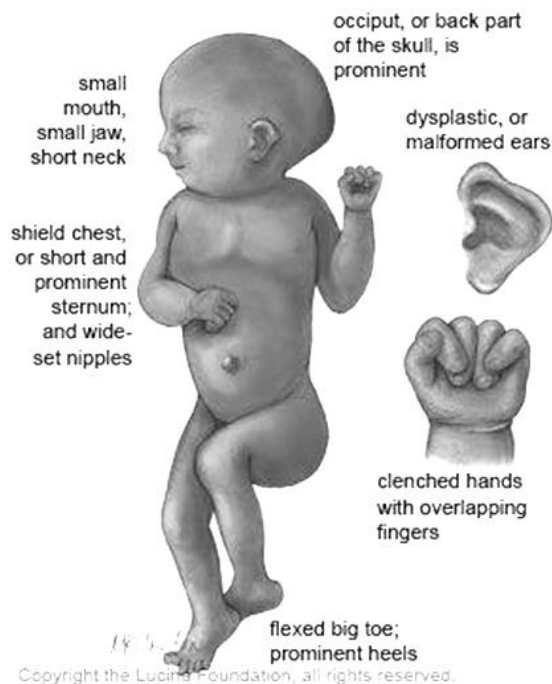


Figure 144. Some signs of Edwards syndrome

Clinical features (Fig. 144):

- Prominent occiput, microcephaly.
- Low-set malformed ears, micrognathia (small jaw), cleft lip and palate, high palate, broad nose.
- Clenched fists, with flexion contracture of the fingers.
- Rocker-bottom feet (convex deformity of the plantar side of the foot).
- Congenital heart defects.
- Malformations of internal organs.
- Severe intellectual disability.

The syndrome is commonly detected in prenatal screening in the second trimester. Genetic testing can be done by cytogenetics tests (e.g. karyotyping) or molecular genetics tests.

There is no effective treatment. The only therapies used are directed toward the specific symptoms developing in each individual case. Sometimes this may include surgical correction of certain abnormalities caused by the disorder.

Prognosis: Only 5–10 % of infants survive past 12 months of age.

Down syndrome (Trisomy 21, karyotype ♀: 47, XX, 21+; ♂: 47, XY, 21+; Fig. 145) is caused by the presence of extra genetic material from the 21st chromosome in some or all of the cells of the body. The extra genetic material disrupts the normal course of development and causes the characteristic features of the syndrome. Only partial trisomy can be inherited (from a parent carrying balanced translocation).

- Full trisomy 21 (~ 95 % of cases) results from meiotic nondisjunction.
- In translocation trisomy 21 (~ 3–4 % of cases) one of the 3 copies of chromosome 21 is attached to another chromosome (usually the 14th one).
- Mosaic trisomy 21 (~ 3–4 % of cases) results from nondisjunction during mitosis.

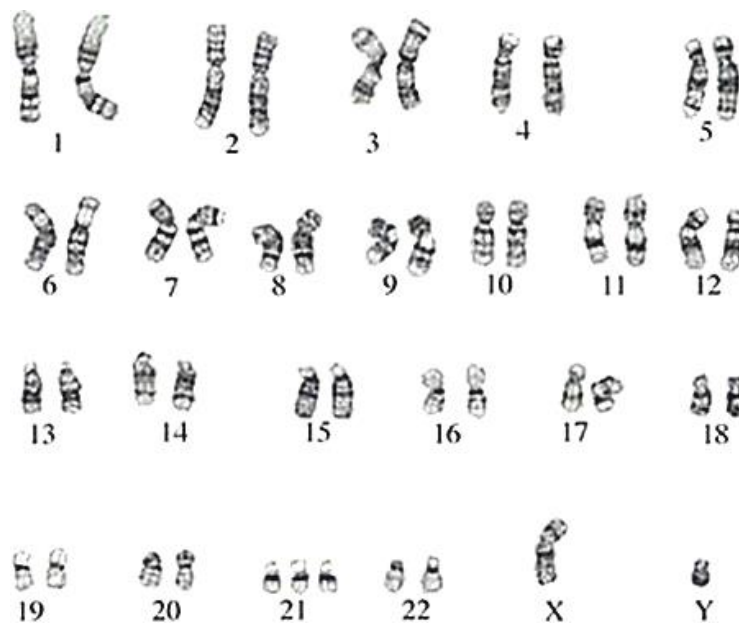


Figure 145. A male karyotype demonstrating trisomy 21

Incidence frequency: ~ 1 : 700 live births. The risk considerably increases with the age of the mother (20 years — 1 : 1500, 40 years — 1 : 100), but not for translocation trisomies.

Clinical features:

- Eyes: upward slanting of the eyelids, epicanthal folds, hypertelorism
Brushfield spots (aggregation of connective tissue in periphery of iris, visible as white or grayish-brown spots, Fig. 146).



Figure 146. Brushfield spots on the iris of a child with Down syndrome

- Mouth: small oral cavity, large tongue, high arched palate, abnormal teeth. Ear anomalies (small, round low-set ears).
- Extremities: single transverse palmar crease, large space between the first and second toe.
- Organ malformations: congenital heart defects in ~ 50 % of cases, defects of the gastrointestinal tract (duodenal atresia/stenosis), and urogenital system.
- Muscle hypotonia.
- Varying levels of intellectual disability (average IQ: 50).
- Diagnosis: prenatal diagnosis (ultrasound, biochemical markers in mother's blood, CVS, amniocentesis with further genetic testing).

There is no treatment to eliminate the disorder. Treatment of the symptom complex and malformations (e.g., heart surgery for cardiac malformations), educational programs, and support are necessary.

Prognosis: reduced average life expectancy (~ 50 years) mostly due to organ malformations (especially cardiac) and increased susceptibility to infections (immunodeficiency).

Cri-du-chat syndrome (♀: 46, XX, 5p⁻; ♂: 46, XY, 5p⁻; Fig. 147) is caused by partial deletion of the short arm of chromosome 5.

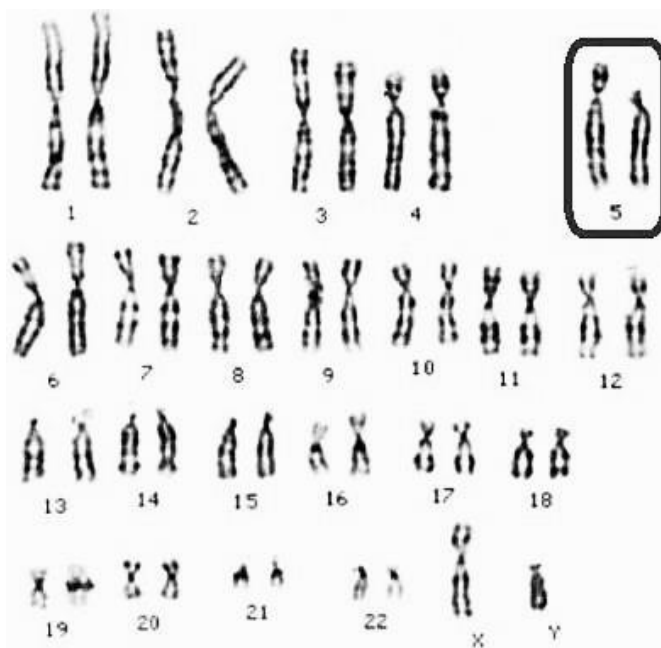


Figure 147. A male karyotype demonstrating deletion in the p arm of the 5th chromosome

The size of the deletion varies among affected individuals; studies suggest that larger deletions tend to result in more severe intellectual disability and developmental delay than smaller deletions. The loss of genetic material from the p arm of chromosome 5 causes the development of specific symptoms of the disorder.

Incidence frequency is 1 in 20,000 to 50,000 newborns ($\text{♀}:\text{♂} = 2 : 1$).

Clinical features:

- Cat-like, high-pitched crying in affected infants.
- Cardiac abnormalities.
- Intellectual disability (moderate to severe).
- Single palmar crease.
- Dysmorphic facial features (e.g., microcephaly, epicanthal folds, broad nasal bridge, downward-slanting palpebral fissures, moon facies, widely spaced eyes).
- Skeletal abnormalities.

Cri-du-chat syndrome is diagnosed in infants and young children on the basis of clinical symptoms and may be confirmed through chromosomal analysis.

Treatment: symptomatic, early psychological and physical assistance.

Prognosis: a normal life expectancy is possible, but depends on the accompanying symptoms and therapeutic assistance.

MITOCHONDRIAL DISEASES

Mitochondrial DNA is especially vulnerable to mutations. One of the reasons is that mitochondria produce large amounts of reactive oxygen species during cellular respiration. They damage various molecules. As a result, mutations in mtDNA occur about ten times more frequently than in nuclear DNA.

Individual cells can have several thousand mitochondria. A zygote receives a large number of organelles through the oocyte, so if the mutation is only in one or a few organelles, it may not affect the cell. During embryonic development, cell division disperses the initial mitochondrial population of the zygote, and these organelles multiply autonomously in the forming cells. Therefore, an embryo and then an adult organism will have cells with normal and mutant organelles. The presence of more than one type of mtDNA in an individual is called heteroplasmy. Mitochondrial diseases are characterized by a correlation between the severity of the course and the number of mutant forms of mtDNA. More than 150 human syndromes are currently known to be inherited in this way. They mostly affect tissues that need a large amount of energy (skeletal muscles, brain).

Examples of mitochondrial diseases:

- MERRF syndrome (myoclonus epilepsy with ragged red fibers). It is a maternally inherited disease. MERRF syndrome is manifested by muscle

coordination disorders, deafness, dementia, and epileptic seizures. Skeletal muscle structure abnormalities are observed, and brain function is impaired. The cells of sick people have a mixture of normal and abnormal mitochondria in different proportions in different patients. Apparently, the presence of only mutant mitochondria is fatal.

- Hereditary Leber Optic Neuropathy (LHON) is maternally inherited and is characterized by sudden bilateral blindness. The average age of vision loss is 27 years, but the onset is highly variable. In more than 50 % of cases, the disease is caused by a mutation in the mtDNA gene encoding the NADH dehydrogenase subunit.

- Kearns–Sayre syndrome (KSS) is characterized by loss of vision, hearing, and heart abnormalities. The genetic basis of the syndrome includes deletions at different mtDNA sites. Many patients develop symptoms in adulthood. The proportion of mutant mtDNA detected is related to the severity of the symptoms.

In addition to mitochondrial diseases, mitochondrial dysfunction is associated with a large number of diseases, including atherosclerosis, Alzheimer's disease, schizophrenia, various types of cancer, and many others. A link between mitochondrial gene damage and aging has been established.

It is estimated that 1 in 5,000 people has mitochondrial DNA abnormalities or is at risk of developing them. Currently, there is no effective treatment for mitochondrial diseases, but an approach to prevent them is being developed: a cytoplasmic transplant from the female donor's cells into the mother's ovum during in vitro fertilization.

POLYGENIC DISEASES

Polygenic (or multifactorial) diseases are diseases determined by the mutual effects of several genes. Because of the complexity of genetic interactions, they develop differently than monogenic diseases. In humans, however, polygenic diseases are more common than monogenic ones. In addition, non-hereditary factors can also play a role in the development of the disease. For example, more than 36 genes have been found to be associated with the risk of developing type 2 diabetes (a disease characterized by elevated blood glucose levels and symptoms such as increased appetite, thirst, and increased urine output). Many of these genes are related to pancreatic beta cell function. Non-genetic factors associated with the onset of diabetes include lifestyle and diet. Other examples of polygenic conditions include coronary heart disease, hypertension, autoimmune diseases, cancers, obesity, and atherosclerosis.

CHAPTER 16 GENETIC COUNSELING

GENETIC COUNSELING, AIMS, AND TASKS

Genetic counseling is a type of specialized care for the population, aimed at preventing the birth of children with congenital malformations and hereditary pathologies. The term “genetic counseling” defines 2 concepts: first, it is a consultation with a geneticist, which ends with the issuance of a medical genetic test report, and second, it is a structural subdivision of the health care institution (genetic center).

The purpose of genetic counseling is to prevent the birth of a child with congenital abnormalities (malformations) and hereditary pathologies. The main tasks of genetic counseling:

1. Retrospective and prospective counseling of families and patients with hereditary or congenital pathologies.
2. Establishment of accurate diagnosis of hereditary disease.
3. Prenatal diagnosis of congenital and hereditary diseases.
4. Assistance to doctors of various specialties in making a diagnosis of the disease, if this requires special methods.
5. Informing the patient and the family about the significance of the information collected about the degree of risk of birth of a child with hereditary pathology and the existing methods of prevention.
6. Making genetic prognosis and assisting parents in making decisions about childbirth.
7. Maintaining a territorial register of families and patients with hereditary and congenital pathologies and their dispensary monitoring.
8. Promotion of medical and genetic knowledge among the population.

Indications for referring a family to genetic counseling:

1. The presence of similar hereditary disorders in several family members.
2. Age of women 35 years and older, men 45 years and older.
3. Two or more spontaneous abortions.
4. Consanguine marriages.
5. Presence of ultrasound markers of malformations or hereditary pathology in the fetus.
6. Exposure to a mutagenic or teratogenic factor before 14 weeks of pregnancy.
7. Birth of a child with hereditary pathology or congenital malformations.
8. Retardation in the mental and physical development of the child.
9. Disorders of puberty.
10. Planning a pregnancy by a woman with phenylketonuria, diabetes, hypothyroidism, etc.

Stages of genetic counseling. The stages of genetic counseling include making a diagnosis, drawing up an opinion, making a prognosis, and giving qualified advice on making a decision.

When a diagnosis of a hereditary or congenital pathology is made, the geneticist determines the scope of the necessary examinations. Genealogical, cytogenetic, biochemical, and other special genetic methods are used to maximize clarification of the diagnosis. It may be necessary to send the patient or his relatives for additional examinations. After all the necessary tests are performed, a medical genetic report is prepared, indicating the diagnosis and the results of the testing, as well as recommendations for medical supervision and treatment.

To make a prognosis and advise a geneticist in a parent's fertility planning decision, it is important to determine the degree of genetic risk and to assess the severity of the medical and social consequences of the presumed abnormality.

Genetic risk refers to the probability of the patient (proband) himself or his relatives developing the abnormality. The average risk of developing a genetically determined abnormality for European populations is 3–5 %, so a risk that does not exceed 5 % is considered low. Genetic risk of up to 10 % is considered elevated to a mild degree, up to 20 % is considered elevated to a medium degree, and over 20 % is considered high. From a genetic point of view, a risk not exceeding a mildly elevated risk can be disregarded and not be considered a contraindication to childbearing even when no prenatal diagnosis of the presumed abnormality is possible. In any case, the family must know the degree of genetic risk to decide whether to plan a pregnancy or terminate it.

The degree of genetic risk does not always correspond to the severity of the expected medical and social consequences of having a child with congenital and hereditary pathology. For example, polydactyly with autosomal dominant inheritance and high genetic risk (at least 50 %) has no significant medical and social consequences. Polydactyly can easily be corrected with appropriate corrective surgery and the person can lead a normal life. Mucopolysaccharidosis, with a 25 % risk in children of heterozygous parents, is a severe disease with significant structural and functional abnormalities in various human organs and is difficult to treat. This disease sometimes manifests itself only in adolescence or even in adulthood and can lead to disability, and the need to constantly take specific medications, which leads to serious social consequences of this pathology. The average frequency of having a child with Down syndrome in European populations of pregnant women under 35 is on average 1 per 1000. This risk is considered very low, at 0.1 %. However, the medical and social consequences of this hereditary pathology are assessed as very severe, since it leads to impaired vitality and intellectual and mental development, and involves the need for special costly programs of education, training, and medical rehabilitation. In the prognosis when planning childbearing,

the geneticist must understandably explain to the patient all the possible risks and consequences, of making a decision.

Not all congenital and hereditary pathologies can be cured. However, different approaches may be available to treat or manage the disease.

Symptomatic treatment. Particular symptoms of all hereditary diseases are treated with medicines: anti-inflammatory drugs for inflammatory processes, painkillers for pains, and sedatives for states of excitation. Surgical treatment is often used in the case of congenital defects such as stenosis and atresia of vessels, polydactyly, heart defects, and defects of the skull.

Pathogenic treatment:

- exchange correction — diet therapy (phenylketonuria, galactosemia);
- metabolic inhibition — suppression of synthesis of the product which is not excreted from the organism (uric acid in the Lesch–Nyhan syndrome);
- substitution therapy — injection of the product that is not produced in the organism (growth hormone in case of dwarfism, insulin in diabetes mellitus).

Etiotropic treatment — elimination of the cause of the disease. The most promising method is gene therapy.

ESTIMATING THE GENETIC RISK

Genetic risk assessment is concerned with determining the probability of having a child with a particular pathology. The probability of an event can be defined as the ratio of the number of times the event occurs to the total number of all possible outcomes. Probability can be expressed in fractions of one (for example, 1 — the event always occurs, 0 — never, 0.25 — in a quarter of all outcomes), as a percentage (100 %, 0 % 25 %, etc.), as a ratio (for example, the probability of having a child with an autosomal recessive disease in a carrier family can be written as a 1 : 3 chance).

It is possible to establish the probability of giving birth to a sick child in a family by analyzing the family tree. Thus, in a family where both parents are carriers of an autosomal recessive disease, the risk of having a sick child is 25 %, a mother who is a carrier of X-linked recessive disease may have a sick son with a 50 % chance, etc.

Calculation of genetic risk. When considering the probability of two different events or outcomes, it is necessary to find out whether they are mutually exclusive or independent, i.e., they cannot or cannot happen simultaneously. Let us consider the difference between such events using the example of blood group inheritance in a family where the parents have genotypes $I^A I^0$ and $I^B I^0$. Their children are equally likely to have genotypes $I^0 I^0$, $I^A I^0$, $I^B I^0$, $I^A I^B$.

$$\begin{array}{l} P: I^A I^0 \times I^B I^0 \\ G: I^A \quad I^0 \quad I^B \quad I^0 \\ F_1: I^0 I^0, I^A I^0, I^B I^0, I^A I^B \end{array}$$

Thus, the children of the described couple can have any blood type with equal probability (25 %). An example of mutually exclusive events is the inheritance of different blood groups — the same child can have only one of them. On the other hand, a child inheriting the I^A allele from one parent does not exclude the possibility of inheriting the I^B allele from another parent, i.e. such events are not mutually exclusive, but independent. If events are mutually exclusive, then the probability that either one or the other will happen is equal to the sum of their probabilities. This is known as the Law of Addition. For example, the occurrence of either a blood type 1 or 2 children in the family described equals $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. If two (or more) events are independent, then the probability that both the first and second happen together is equal to the product of their probabilities. This is known as the Law of Multiplication. For example, in the family discussed above, the probability of a parent with the genotype $IAI0$ passing the IA allele to his or her child, rather than $I0$, is $\frac{1}{2}$. The same is true for inheriting the IB allele from the second parent. Therefore, the probability of getting the $IAIB$ genotype is $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$. Another example: the probability of having a healthy son from a mother who is a carrier of hemophilia is $\frac{1}{2}$, and the probability of having two healthy sons is $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$.

BAYES' THEOREM

Consider a pedigree in which the proband is a woman whose mother is a carrier of hemophilia (an X-linked recessive disease, Fig. 148).

In this situation, the proband woman could have inherited either the X^H or X^h allele from her mother, i.e. there should be a 50 % chance that she is a carrier of the disease (Fig. 149).

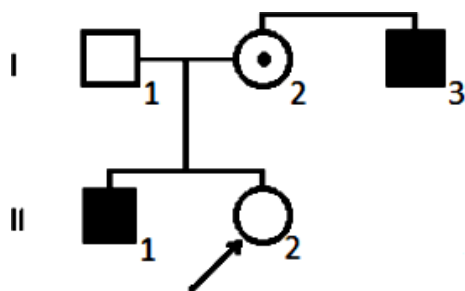


Figure 148. The case of X-linked recessive inheritance (II generations)

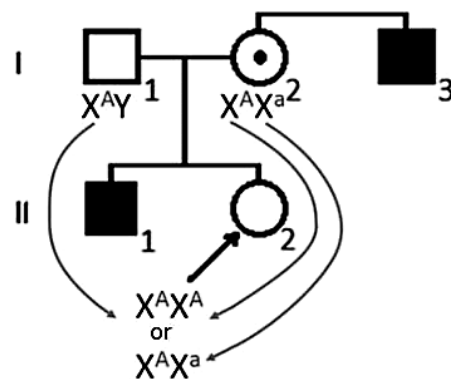


Figure 149. The illustration of X^h allele inheritance

If a woman is a carrier, then, when a son is born, the probability that he will be healthy is $\frac{1}{2}$. However, the probability of a carrier having two healthy sons alone is $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$; three is $\frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{8}$. Thus, in this situation, the probability that the proband is a carrier (initially 50 %) decreases with the birth of each healthy son. The birth of a healthy son does not exclude

the probability of being a carrier, but if the woman in question had a large number of healthy sons and they were all healthy, it would be highly probable to conclude that she is not a carrier of hemophilia.

The “initial” probability of an event, i.e., obtained without taking into account observational data is called a prior probability. In the example described, it is the probability of a proband 50 % being a carrier of hemophilia (the probability of inheriting an X chromosome with a mutation).

The probability of an event corrected for observational data is called a posterior probability. In the example under consideration, this is the probability of being a carrier in the presence of some number of healthy sons (and no sick sons, since this would mean that the proband is a carrier of the mutant allele).

To calculate the posterior probability in such cases, Bayes’ theorem (alternatively Bayes’ law or Bayes’ rule) is used. It can be formulated as follows: the probability of a hypothesis after receiving new empirical data is equal to the product of its probability before receiving this data and the probability of receiving this data due to this hypothesis, divided by the probability of receiving the same data in the light of all possible different hypotheses. Let’s go back to the example of the woman who potentially carries the mutation that causes hemophilia (Fig. 150).

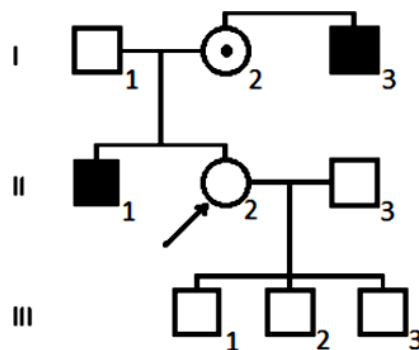


Figure 150. The case of X-linked recessive inheritance (III generations)

The birth of three healthy sons (in the absence of affected sons) is possible both in the case of the carrier (prior probability is $\frac{1}{2}$) and is $\frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{8}$, and in the case of non-carrier (a priori probability is also $\frac{1}{2}$), and is 1 (i.e. 100 %).

Thus, the probability of having three healthy children in the light of both hypotheses (both carrier and non-carrier) is $\frac{1}{2} * \frac{1}{8} + \frac{1}{2} * 1$, i.e., $\frac{9}{16}$. To calculate the posterior probability that the proband is a carrier, divide the probability of having three healthy children in the case of carrier multiplied by the probability of carrier ($\frac{1}{8} * \frac{1}{2}$) by the probability of having a healthy child in the light of both hypotheses ($\frac{9}{16}$):

$$P = (\frac{1}{2} * \frac{1}{8}) / (\frac{1}{2} * \frac{1}{8} + \frac{1}{2} * 1) = (\frac{1}{16}) / (\frac{9}{16}) = \frac{1}{9} \text{ (Fig. 151).}$$

Thus, it can be concluded that the probability of carrying the hemophilia-causing mutation considered by the female proband is not $\frac{1}{2}$, as follows from the analysis of the pedigree, but $\frac{1}{9}$.

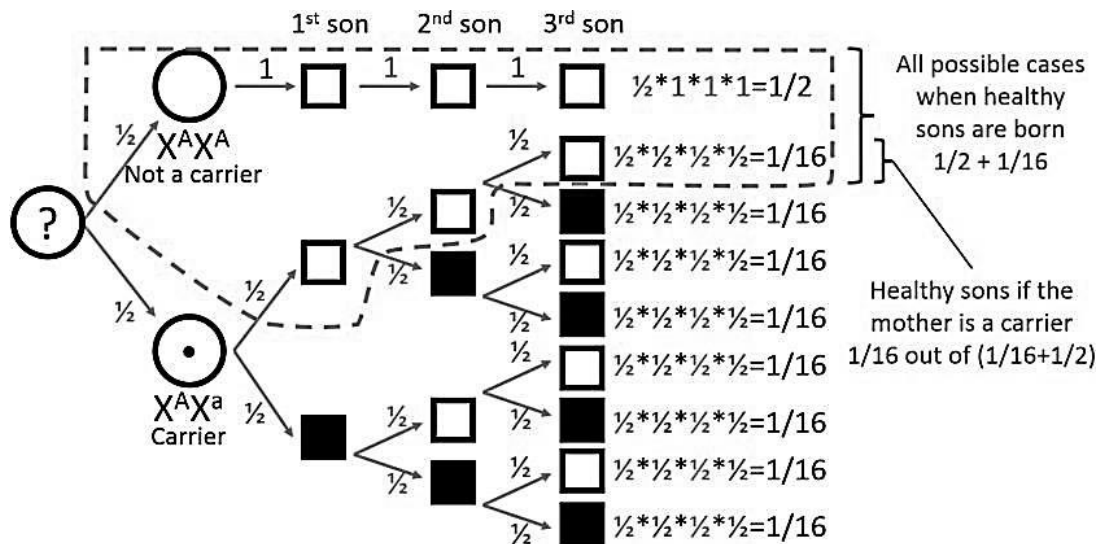


Figure 151. Illustration of the posterior probability calculation

PRENATAL TESTING

An important aspect influencing the prognosis and the physician's advice is the perspective and effectiveness of the prenatal diagnostic tests available at the genetic center. Advances in this field make it possible to plan births in families with a high risk of inheriting severe pathologies (Down syndrome, mucopolysaccharidosis, hemophilia, cystic fibrosis, etc.) since these diseases can be detected by prenatal diagnostic methods.

An example of prenatal testing based on the practice of the Republic of Belarus. The first stage is ultrasound and biochemical screening in the *first trimester* of pregnancy. First-trimester screening (11–14 weeks) includes fetal ultrasound and blood marker compounds (alpha-fetoprotein (AFP), Human chorionic gonadotropin (hCG), and estriol free (E3). Fetal anatomy, nuchal translucency (NT), nasal bones, fronto-occipital size, and blood flow in the umbilical cord and uterine arteries are assessed. If ultrasound signs of fetal malformations, markers of chromosomal abnormalities, or deviations in biochemical markers are detected, the pregnant woman is referred for consultation to a geneticist for additional studies, which include ultrasound on an expert-level machine and invasive prenatal diagnosis: chorionic villus sampling, amniocentesis, cordocentesis, and others.

Second-trimester screening (18–21 weeks). An ultrasound is performed to exclude late congenital fetal abnormalities and Quad Screen Test (four markers in the maternal serum: AFP, E3, hCG, and inhibin A (corpus luteum hormone). If any abnormalities are detected, the pregnant woman is referred to a geneticist for additional testing.

Third-trimester screening (32–35 weeks). At this time, only an ultrasound examination is performed to detect fetal growth retardation and malformations with a late manifestation.

At the current level of development of prenatal diagnosis, it is possible to diagnose all chromosomal diseases, most birth defects, and inborn errors of metabolism. Some of them can be diagnosed at almost any gestational age (chromosomal diseases), some after the 12th week (limb reduction defects, atresia, anencephaly), and some only in the second half of pregnancy (heart and kidney defects). Indirect and direct methods of prenatal diagnosis are currently used. With indirect methods, the pregnant woman is examined (obstetric-gynecological methods, blood serum for alpha-fetoprotein). With direct methods, the fetus is examined. Direct noninvasive (without surgery) methods include ultrasonography. Direct invasive methods (with disruption of tissue integrity) include chorion biopsy, amniocentesis, cordocentesis, and fetoscopy.

Indications for invasive prenatal testing are:

1. Increased risk of chromosomal abnormalities in the fetus based on first-trimester screening.
2. Ultrasound markers of fetal chromosomal abnormalities detected in the first or second trimester of pregnancy.
3. Fetal congenital abnormalities (malformations) detected by first or second-trimester ultrasound screening.
4. Age of the pregnant woman is over 35 years old.
5. Presence of chromosomal rearrangement in one of the spouses.
6. Presence of a child (fetus) with a chromosomal abnormality in the anamnesis.
7. Increased risk of having a child with a monogenic disease, for which specific methods of prenatal diagnosis have been developed.

Alpha-fetoprotein (AFP) is a protein produced in the liver of a developing fetus. During development of the fetus, some AFP passes through the placenta and into the mother's blood. An **AFP test** measures the level of AFP in pregnant women and is generally done between 15 and 20 weeks of pregnancy. Too much or too little AFP in a mother's blood may be a sign of a birth defect or other condition. These include:

- A neural tube defect, a serious condition that causes abnormal development of a developing baby's brain and/or spine.
- Down syndrome and other chromosome disorders.
- Twins or multiple births, because more than one baby is producing AFP.
- Miscalculation of the due date, because AFP levels change during pregnancy.

A pregnancy ultrasound is an imaging test that uses sound waves to create a picture of how a fetus is developing in the womb. It is also used to check the female pelvic organs during pregnancy.

An ultrasound may be done to determine if there is a problem with the pregnancy, how far along the pregnancy is, or to take measurements and screen for potential problems.

A pregnancy ultrasound may be done during the first 12 weeks of pregnancy to:

- Confirm a normal pregnancy.
- Determine the baby's age.
- Look for problems, such as ectopic pregnancies or the chances for a miscarriage.
- Determine the baby's heart rate.
- Look for multiple pregnancies (such as twins and triplets).
- Identify problems of the placenta, uterus, cervix, and ovaries
- Look for findings that might indicate an increased risk for Down syndrome.

A pregnancy ultrasound may also be done in the second and third trimesters to:

- Determine the baby's age, growth, position, and sometimes sex.
- Identify any problems with how the fetus is developing.
- Look for twins or triplets. Look at the placenta, amniotic fluid, and pelvis.

Chorionic villus sampling (CVS) is a test for pregnant women that checks chorionic villi, tiny finger-like growths found in the placenta. The test looks for abnormalities in the baby's chromosomes. Chorionic villi contain the same chromosomes as an unborn baby. So, a CVS test can show if a baby has an extra chromosome or a missing or damaged chromosome.

CVS testing can be done early in pregnancy, between the 10th and 13th week. There are two types of CVS tests:

Transabdominal. The sample is taken through the abdominal wall. A long, thin needle is inserted through the abdomen and uterus and into the placenta to withdraw a sample of tissue from the placenta.

Transcervical. The sample is taken through the cervix. A thin tube is inserted through the vagina and cervix and up to the placenta to suck in a sample of tissue from the placenta.

Amniocentesis is a test for pregnant women that looks at a sample of amniotic fluid. Amniotic fluid is a pale, yellow liquid that surrounds and protects an unborn baby throughout pregnancy. The fluid contains cells that provide important information about the unborn baby's health.

Amniocentesis is used to diagnose certain health problems in an unborn baby. These include:

- Genetic disorders, which are often caused by mutations in certain genes. These include cystic fibrosis and Tay–Sachs disease.
- Chromosome disorders. As well as sex chromosome anomalies.
- Sex determination.
- A neural tube defect.
- The test may also be used to check your baby's lung development.

The test is usually done between the 15th and 20th weeks of pregnancy. It is sometimes done later in pregnancy to check the baby's lung development or diagnose certain infections. Technique:

- Numbing medicine may be applied to the abdomen.
- An ultrasound device is moved over the abdomen.
- Using the ultrasound images as a guide, a thin needle will be inserted through the abdominal wall and withdraw a small amount of amniotic fluid.

Cordocentesis — also known as percutaneous umbilical blood sampling — is a diagnostic prenatal test in which a sample of the baby's blood is removed from the umbilical cord for testing. Cordocentesis, which is usually done after week 18 of pregnancy, can be used to detect certain genetic disorders, blood conditions, and infections. Cordocentesis can also be used to deliver blood and medication to a baby through the umbilical cord.

Guided by ultrasound, a thin, hollow needle is inserted through the abdominal wall and into the uterus. A small amount of blood from the vein in the umbilical cord will be withdrawn into a syringe, and the needle will be removed.

Fetoscopy is a procedure during pregnancy to allow surgical access to the fetus, the amniotic cavity, the umbilical cord, and the fetal side of the placenta. A small (3–4 mm) incision is made in the abdomen, and an endoscope is inserted through the abdominal wall and uterus into the amniotic cavity. If needed, collect tissue samples or perform surgery.

Fetoscopy is usually performed in the second or third trimester of pregnancy.

ETHICAL ISSUES IN GENETIC COUNSELING

Ethical problems in the practical activity of a geneticist are more often related to the ambiguous attitude towards abortion after prenatal diagnosis of fetal pathology with the notions of necessary confidentiality of genetic information, accessibility of genetic information to relatives, spouses, and third parties. Also, ethical problems arise in the testing of diseases with delayed manifestation.

One of the main principles of genetic counseling, supported by the majority of geneticists in different countries, is the principle of non-directiveness. Non-directiveness includes two main components: the first one is to provide the patient with accurate, complete, and unbiased information, which the counselor can understand and use when making a decision; the second one is psychological support by a geneticist, a sympathetic attitude towards the patient, assistance in developing their own decision, respect for the decision made by the patient. In non-directive counseling, a geneticist should, as much as possible, support the patient's own decision. This does not mean that the doctor has no right to assess the situation, but even if they are sure in their

assessment, the doctor does not impose it on the patient or their family. The patient decides for themselves whether or not to undergo testing, which tests to perform and which not to perform, what to do with the genetic information obtained — to disclose it to their spouse and relatives or not to disclose it, and the choice to carry the fetus with abnormalities to term or to have an abortion based on the prenatal diagnosis results.

Directiveness, on the other hand, is the intentional attempt of the consultant to push the patient toward a certain decision, for example, explicitly persuading parents to make a certain reproductive decision, emphasizing the negative aspects, so that they are inclined to terminate the pregnancy without the direct recommendation.

The WHO expert panel recommends that genetic counseling can be directive only when a prenatally diagnosed disease can be cured. In the United States, Canada, and the United Kingdom, where there is an official ideology of privacy and a tradition of respect for privacy, non-directive counseling is preferred; in European countries, geneticists are more prone to a directive approach.

The desire for non-directive genetic counseling is now a common attitude throughout the world, but the difficulties and limitations associated with its implementation in practice are also recognized in the fundamental bioethical documents, in particular the Proposed International Guidelines on Ethical Issues in Medical Genetics and Genetic Services.

International and national ethical regulations reserve to the mother and the family the right to make an autonomous choice to maintain or terminate a pregnancy with a fetus with a pathology, subject to the obligation to provide truthful and comprehensive information in an accessible form about all possible consequences. Reproductive decisions should be the prerogative of those who will be directly responsible for the biological and social aspects of childbirth and upbringing. In cases where a couple cannot agree, the mother should have the last word.

The preimplantation genetic diagnosis implies the possibility to detect embryo abnormalities and perform embryo selection before implantation into the uterus before pregnancy begins. Currently, in developed countries, thanks to genetic diagnosis, embryos can be tested for the presence of more than 20 pathological alleles that provoke monogenic pathology, as well as for chromosomal aberrations. Therefore, despite the apparent similarity of both types of diagnosis, the response to the results of preimplantation genetic diagnosis is psychologically much less painful for the couple than to the results of prenatal diagnosis.

There are objections to preimplantation and prenatal diagnosis because this violates the principle of the equal worth of all people, including the sick. The opposing side argues that one should not confuse a particular genetic disease with the person suffering from that disease. The fact that a disease is

evaluated negatively, that it is legitimately sought to be avoided, does not necessarily mean a negative evaluation of the person suffering from it. There is no conflict of interest in this approach between trying to avoid the disease, reducing its prevalence, and helping those people who suffer from it.

THE PROBLEM OF RESPECTING THE CONFIDENTIALITY OF GENETIC INFORMATION

According to the rules of genetic counseling, information about a person's genetic status can only be disclosed to them, to their guardian or another legal representative, and their treating physicians.

The need to maintain the confidentiality of genetic information is consistent with one of the basic principles of bioethics — do not harm. The rule of confidentiality permits the disclosure of information obtained in genetic research only with the consent of the patient, and only “for compelling reasons in the public interest, clearly specified in domestic law consistent with the international law of human rights” without the patient's consent. The problem of disposing of information about a patient's genetic status in a situation in which family members, relatives, or spouses may have an interest in it raises serious ethical debates. Of course, it is unethical to disclose information about a person to others without their consent, especially since this information can be harmful to the patient, leading to discrimination in the family and at work. But is it ethical to withhold genetic information from a patient's blood relatives, who may also have a pathological allele and risk developing a hereditary disease in themselves and their children? In addition to blood relatives, the spouse is interested in genetic information about the patient, especially if the couple is planning to have children. However, this information can serve to the patient's detriment, such as destroying his or her marriage.

LITERATURE

1. *Campbell Biology* / L. A. Urry [et al.]. – 12th ed. – NJ : Pearson Education, 2021. – 1493 p.
2. *Biology* / P. Raven [et al.]. – 12th ed. – NY : McGraw-Hill Education, 2020. – 1472 p.
3. *Ченцов, Ю. С.* Введение в клеточную биологию : учеб. для вузов / Ю. С. Ченцов. – М. : Академкнига, 2004.
4. *Molecular Cell Biology* / H. Lodish [et al.]. – 8th ed. – NY : W. H. Freeman and Company, 2016. – 1278 p.
5. *Медицинская биология и общая генетика* / Р. Г. Заяц [и др.]. – Минск : Вышэйшая школа, 2017. – 481 с.
6. *Concepts of genetics* / W. S. Klug [et al.]. – 12th ed. – NY : Pearson Education, 2019. – 867 p.
7. *Патрушев, Л. И.* Экспрессия генов / Л. И. Патрушев. – М. : Наука, 2000.
8. *Инге-Вечтомов, С. Г.* Генетика с основами селекции / С. Г. Инге-Вечтомов. – Изд-во Н-Л, 2010. – 720 с.
9. *Molecular biology of the cell* / B. Alberts [et al.]. – 6th ed. – New York, 2015.

CONTENTS

Chapter 1. Medical Biology and its role in medical education. The subject, tasks, and methods of Cytology	3
Chapter 2. Structural and functional organization of the cell.....	15
Chapter 3. Structure of DNA	29
Chapter 4. The cell cycle	44
Chapter 5. The flow of genetic information in the cell	58
Chapter 6. Regulation of gene expression	74
Chapter 7. Genomics. Techniques of molecular genetics	91
Chapter 8. Genetic engineering	106
Chapter 9. Genetically engineered products.....	119
Chapter 10. Basic laws of inheritance	138
Chapter 11. Genetic linkage. Biology and genetics of sex.....	151
Chapter 12. Variation. Mutagenesis. Carcinogenesis.....	164
Chapter 13. Population genetics	173
Chapter 14. Human genetics and medical genetics	184
Chapter 15. Hereditary disorders.....	197
Chapter 16. Genetic counseling.....	213
Literature.....	223