



DEBRECENI EGYETEM

University of Debrecen  
Faculty of Agricultural and Food Science and  
Environmental Management  
Institute of Food Science

# BACTERIAL AND YEAST MOLECULAR BIOLOGY

---

Erzsébet Sándor, Ph.D. – Anikó Szojka

2015

ISBN 978-963-473-830-5

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME

**SZÉCHENYI** 



MAGYARORSZÁG  
KORMÁNYA

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



# Contents

1	DESCRIPTION OF GRAM-NEGATIVE AND GRAM-POSITIVE CELLS, AND YEAST CELL STRUCTURE .....	3
2	DESCRIPTION OF INFORMATIONAL MACROMOLECULES , GENES AND GENE EXPRESSION .....	7
3	DNA STRUCTURE AND REPLICATION, TRANSCRIPTION AND PROTEIN SYNTHESIS .....	9
4	EUKARYOTIC GENETICS AND MOLECULAR BIOLOGY .....	12
5	PRINCIPLES OF BACTERIAL GENETICS AND PROKARYOTIC GENOMES .....	15
6-7	REGULATION OF GENE EXPRESSION I-II .....	18
8	BACTERIAL GENE MANIPULATION.....	23
9	TOOLS OF GENETIC ENGINEERING TECHNIQUES I (RESTRICTION AND MODIFICATION ENZYMES, ESSENTIAL MOLECULAR CLONING) .....	26
10	TOOLS AND TECHNIQUES OF GENETIC ENGINEERING II (PLASMIDS, SHUTTLE AND EXPRESSION VECTORS, PLASMIDS EXPORT, OTHER VECTORS: YAC, BAC).....	29
11	TOOLS AND TECHNIQUES OF GENETIC ENGINEERING III (NUCLEIC ACID HYBRIDIZATION AND SOUTHERN BLOT).....	32
12	SEQUENCING, SYNTHESIS AND AMPLIFICATION OF DNA.....	36
13	GENE MANIPULATIONS: SOME EXAMPLES OF APPLICATIONS IN BIOTECHNOLOGY AND BASIC RESEARCH .....	40
14	METABOLIC ENGINEERING OF MICROORGANISMS FOR THE PRODUCTION OF MOLECULES OF INTEREST FROM BIOMASS	42
15	GENOME FUNCTION AND REGULATION (MICROARRAYS AND TRANSCRIPTOME, PROTEOMICS, METABOLOMICS).....	44





## 1 DESCRIPTION OF GRAM-NEGATIVE AND GRAM-POSITIVE CELLS, AND YEAST CELL STRUCTURE

---

The cell is the fundamental unit of life. Cells are isolated from the surrounding environment by a semipermeable membrane. That means that the cell is an open, dynamic structure, with exchanging materials and communicating with its environments.

All cells have similar properties and containing several common components, like cytoplasmic membrane, cytoplasm and chromosome(s). The cytoplasm is an aqueous mixture of macromolecules, small organic molecules, and inorganic ions, which is surrounded by cytoplasmic membrane. It also contains ribosomes, the cell's protein-synthesizing structures. All cells possess one or more chromosomes containing the majority of its genetic information.

Regarding the internal structure of cells two distinct type can be separated: (i) prokaryote and (ii) eukaryote. In terms of cell size, a typical rod-shaped prokaryote is 1–5 micrometres long and about 1 micrometre wide. The range of sizes in eukaryotic cells is quite large. Eukaryotic cells are usually much larger and structurally more complex than prokaryotic cells. Eukaryotic cells are known with diameters as small as 0.8 micrometres or as large as several hundred micrometres.

Eukaryotic organisms are the protists (algae and protozoa), the fungi and slime moulds, the plants and animals. A major property of eukaryotic cells is the presence of membrane-enclosed structures in the cytoplasm called organelles: the nucleus, the mitochondria and, in photosynthetic cells, chloroplasts. In eukaryotic cell, the cell's genome is arranged in linear DNA molecules within the membrane-enclosed nucleus. Mitochondria and chloroplasts are dedicated to energy conservation and carry out respiration and photosynthesis, respectively. The key processes of genetic information flow are partitioned in the eukaryotic cells: DNA replication and transcription (RNA synthesis) occur in the nucleus while translation (protein synthesis) occurs in the cytoplasm.

In contrast to eukaryotic cells, prokaryotic cells have a simpler internal structure without organelles. However, prokaryotes differ from eukaryotes in many other ways as well. In most prokaryotic cells, DNA is present in circular molecule called the chromosome, usually only one. However it must be noted, that few prokaryotes have linear chromosome, and sometimes two or even three. The chromosome aggregates within the cell to form a mass called the nucleoid. As prokaryotes have only one chromosome, they are typically genetically haploid (contain only one copy of each gene). Many prokaryotes also contain one or more small circles of DNA, called plasmids. The transcription and the translation are both in the cytoplasm. That means they are not separated neither in space, nor in time in the prokaryotic cells. Moreover cytoplasmic membrane is employed in the energy-conservation reactions in most prokaryotes.

The prokaryotic cell is simpler than the eukaryotic cell at every level, with one exception: the cell envelope is more complex. The cell envelope refers the layers that bound the prokaryotic cells. The structure and the organization of the cell envelope differ in gram-positive and gram-

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



negative bacteria, which define these two major groups of bacterial species. The gram-negative cell envelope is chemically complex and consists of several layers, whereas the gram-positive cell wall is typically much thicker and consists primarily of a single type of molecule. The distinction between gram-positive and gram-negative bacteria is based on the Gram stain reaction.

The envelope of gram-positive cells is relatively simple, consisting of just three layers: (i) the cytoplasmic membrane, (ii) a thick peptidoglycan layer, and sometimes (iii) a variable outer layer, called capsule.

The gram-negative cell envelope is a complex, multilayer structure. The (i) cytoplasmic membrane (also called inner membrane) is surrounded by a (ii) thin peptidoglycan. The (iii) complex outer membrane is anchored to the peptidoglycan. The space between the inner and the outer membrane is called (iv) periplasmic space. As an outermost layer, (v) capsule may also present.

The bacterial cytoplasmic membrane is composed of phospholipids and proteins. Selective transport and permeability is one of the major functions of the cytoplasmic membrane. Specific membrane-integrated proteins (permeases) in the cytoplasmic membrane can facilitate the diffusion of specific solutes, or catalyse energy-dependent active transport. In gram-negative bacteria specific binding in the periplasmic space proteins may facilitate the transport of many nutrients with binding the nutrient, than passed to a transport carrier protein of the inner membrane.

The rigid peptidoglycan layer of bacterial envelope is primarily responsible for the strength of the wall. In gram-positive bacteria, as much as 90% of the wall is peptidoglycan. In gram-negative bacteria, additional layers are present outside this rigid layer. The peptidoglycan is a polysaccharide composed of two sugar derivatives: N-acetylglucosamine and N-acetylmuramic acid. A short oligopeptide chain is connected to the N-acetylmuramic acid, containing L-alanine, D-alanine, D-glutamic acid, and either lysine or the structurally similar amino acid analogue, diaminopimelic acid (DAP) amino acids. These tetraptide chains are connected through cross-links of amino acids. The covalent glycosidic bonds connecting the sugars in the glycan strands providing rigidity to the structure in one direction. The cross-linking of peptidoglycan provides stability of the peptidoglycane in all direction.

In gram-negative bacteria, the amino group of DAP of one glycan chain to the carboxyl group of the terminal D-alanine on the adjacent glycan chain creates cross-linkage in the peptidoglycan. In gram-positive bacteria, cross-linkage may occur through a short peptide interbridge. The nature and numbers of amino acids in the interbridge is varying from species to species.

Peptidoglycan can be destroyed by certain agents, like lysozyme. This enzyme cleaves the  $\beta$ -1,4-glycosidic bonds between N-acetylglucosamine and N-acetylmuramic acid in peptidoglycan, consequently the wall weaken, resulting cell lysis. Penicillin however acts with preventing the biosynthesis of peptidoglycan, leading to osmotic lysis as well.

Many gram-positive bacteria have acidic components called teichoic acids embedded in their cell wall. It is a polyalcohol, connected by phosphate esters and typically contains sugars or D-alanine. Teichoic acids are covalently bonded to muramic acid in the wall peptidoglycan. Certain teichoic acids are covalently bound to membrane lipids, and these are called lipoteichoic acids.

The peptidoglycan content of gram-negative envelope only about 10%, but they possess a second lipid bilayer called outer membrane. This

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



layer is not constructed solely of phospholipid and protein, like cytoplasmic membrane, but also contains polysaccharide. The lipid and polysaccharide are linked in the outer membrane to form a complex (lipopolysaccharide - LPS). The polysaccharide portion of LPS consists of two components, the core polysaccharide and the O-polysaccharide. The outer half of the outer membrane bilayer is mainly, while lipoprotein is present on the inner half of the outer membrane, along with usual phospholipids. Lipoprotein is anchoring the outer membrane to peptidoglycan. Several porins are embedded in the outer membrane providing the outer membrane of gram-negative bacteria to be relatively permeable to small molecules. Porins are transmembrane proteins forming channels for the entrance and exit of solutes. Several porins are known, including both specific and nonspecific classes.

Periplasm is located between the outer surface of the cytoplasmic membrane and the inner surface of the outer membrane. It contains high concentration of several different classes of proteins, including hydrolytic enzymes for initial degradation of food molecules; binding proteins participating in the transport of substrates; and chemoreceptors involving in the chemotaxis.

The differences in the Gram stain reaction of gram-positive and gram-negative Bacteria is considered based on the structural differences between their cell walls. The insoluble crystal violet–iodine complex forms inside the cell during the first steps of the Gram stain. Gram positive bacteria have very thick cell walls (mainly of peptidoglycan) and these become dehydrated by the alcohol, applying as a differentiation step during the Gram staining. Consequently alcohol prevents the insoluble crystal violet–iodine complex to escape with closing the pores in the walls. However the complex is extracted by alcohol from gram-negative bacteria, as alcohol readily penetrates the lipid-rich outer membrane and extracts the crystal violet–iodine complex from the cell. After alcohol treatment, gram-negative cells are nearly invisible unless the application of a second dye, which is a standard procedure in the Gram stain.

Yeast cells are eukaryotic fungi; possess the characteristics of eukaryotic cells. All eukaryotes contain a membrane-enclosed nucleus and, several other organelles. Yeast cells also have mitochondria Golgi complex, peroxisomes, vacuoles, endoplasmic reticula, microtubules and microfilaments. Moreover cell wall is a structural component of yeast cells. Cell wall represents 26 - 32 % of the dry weight of *Saccharomyces* and other kinds of yeast. The yeast cell wall (similarly to the bacterial one) has the function of (i) limiting the water influx from outside, thereby stabilization of internal osmotic conditions; (ii) protection against physical stress and mechanical damage; (iii) maintenance of cell shape, which is a precondition for morphogenesis; (iv) serving as a scaffold for different proteins.

The yeast cell wall has an inner layer of load-bearing polysaccharides (mainly moderately branched 1,3- $\beta$ -glucan), acting as a scaffold for a protective outer layer of mannoproteins that extend into the medium. At the external face of the 1,3- $\beta$ -glucan network, highly branched (and thus water-soluble) 1,6- $\beta$ -glucan chains are found, which in turn may be connected to a glycosylphosphatidylinositol (GPI)-modified mannoproteins. Chitin chains may also attach, either to 1,3- $\beta$ -glucan or to 1,6- $\beta$ -glucan chains, but only after cytokinesis at the ring in the neck of the mother cell, in the primary septum and in bud scars.

The nucleus contains the chromosomes of the yeast cell, where DNA within is wound around proteins called histones. The nucleus is enclosed by a two membranes. The inner membrane is a simple sac; the outer membrane is continuous with the endoplasmic reticulum. The nuclear membranes

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



contain pores, which allow the import and export other proteins and nucleic acids into and out of the nucleus in a process called nuclear transport. The nucleolus is within the nucleus. This is the site of ribosomal RNA (rRNA) synthesis.

Respiration and oxidative phosphorylation are localized in mitochondria (singular, mitochondrion). Mitochondria are of bacterial dimensions. A yeast cell contains several mitochondria. The mitochondrion is enclosed by a double membrane system. The invagination of the inner membrane forms cristae, bearing the enzymes for respiration and ATP production. Cristae also contain specific transport proteins that regulate the passage of metabolites, in particular ATP into and out of the matrix of the mitochondrion. The matrix contains enzymes for the oxidation of organic compounds—in particular, enzymes of the citric acid cycle.

The endoplasmic reticulum (ER) is a network of membranes. There are types of endoplasmic reticulum: (i) rough, which contains attached ribosomes and continuous with the nuclear membrane, (II) and smooth. The endoplasmic reticulum (ER) in *Saccharomyces cerevisiae* consists of a reticulum underlying the plasma membrane (cortical ER) and ER associated with the nuclear envelope (nuclear ER).

The Golgi complex consists of a stack of membranes distinct from the ER. The products of the ER are chemically modified and sorted into those destined for secretion in the Golgi complex. Its morphology varies among yeast species: in the budding yeast *S. cerevisiae* it exists as a system of dispersed cisternae, whereas in other species (*Pichia pastoris* and *Schizosaccharomyces pombe*) are organized into stacks.

Lysosomes are membrane-enclosed compartments, containing various digestive enzymes that hydrolyze macromolecules (such as proteins, fats, and polysaccharides), and damaged cellular components and recycling these materials for new biosynthesis. The lysosome allows the cell's lytic activities to be partitioned away from the cytoplasm. Fungal vacuoles are acidic organelles. Similarly to mammalian lysosomes and plant vacuoles, they have degradative and storage capabilities.

The peroxisome is also a membrane-enclosed, specialized metabolic compartment. Its function is to oxidize various compounds (alcohols, long-chain fatty acids), and breaking them down into smaller molecules. They also oxidize toxic compounds in the cell. The enzymes of the peroxisome transfer hydrogen from these compounds to  $O_2$ , producing hydrogen peroxide ( $H_2O_2$ ) as a by-product, which is then degraded to  $H_2O$  and  $O_2$  by catalase enzyme.

Eukaryotic cells bear cytoskeleton system, an internal structural network, comes from proteins that form filamentous structures called microtubules, microfilaments, and intermediate filaments. Microtubules are tubes are composed of the  $\alpha$ -tubulin and  $\beta$ -tubulin, being important in chromosome movement during cell division, and in movement of organelles in yeast cells. Microfilaments are smaller filaments, and are polymers of two intertwined strands of the protein actin. Microfilaments function in maintaining and changing cell shape, during cell division.





## 2 DESCRIPTION OF INFORMATIONAL MACROMOLECULES , GENES AND GENE EXPRESSION

---

Cells can be considered as a kind of chemical machines, as they are able to transform macromolecules into new cells. Those macromolecules are (i) polysaccharides, (ii) lipids, (iii) proteins, and (iv) nucleic acids. The later ones play important roles in storing, processing, and using genetic information.

The functional unit of genetic information is the gene. Genes are universal, as all cellular organisms, and viruses contain it. Genes are located either on chromosomes or on other genetic elements. The genetic information is carried by nucleic acids. Deoxyribonucleic acid (DNA) carries the genetic information in cells, and ribonucleic acid (RNA) is the intermediary molecule converting the information of DNA into defined amino acid sequences in proteins. The sequence of monomers is the code in the nucleic acid macromolecules the genetic in the nucleic acids. The sequence of the nucleic acids determines the sequence of monomers in proteins, that means both nucleic acids and proteins are also informational macromolecules, while lipids and polysaccharides are not.

DNA and RNA are polynucleotides. A nucleotide has three components (i) a pentose sugar, (ii) a nitrogen base, and (iii) a phosphate molecule. The general structure of nucleotides of both DNA, and RNA is very similar, but there are differences between them. DNA contains deoxyribose, while ribose is in RNA. The nitrogen bases are either purines (adenine and guanine) or pyrimidines (thymine, cytosine, and uracil). Purines contain two fused heterocyclic rings, and pyrimidines contain a single six-membered heterocyclic ring. Guanine, adenine, and cytosine are present in both DNA and RNA. The other difference between RNA and DNA is that thymine is present only in DNA and uracil is present only in RNA. A nitrogen base attached to its sugar in nucleosides without phosphate group. The bases are attached to the pentose sugar by a glycosidic linkage between C1 of the sugar and a N1 (in pyrimidine) between C1 of the sugar and a N9 (in purine bases).

Nucleotides are covalently bonded between the 3'- (3 prime) carbon of one sugar and the 5'-carbon of the next sugar both forming a backbone with alternating sugar and phosphate. That means phosphate connects two sugars by phosphodiester bond (ester linkage) in the DNA and the RNA molecules.

Nucleic acids have primary and secondary structure, important in the coding and expressing of genetic information. The sequence of nucleotides creates the primary structure of nucleic acids (DNA and RNA). That means, the genetic information is coded in the sequence of bases. Secondary structure of RNA molecules shows the folding of single-stranded RNA, where complementary base pairing is possible.

DNA has a special structure with double-stranded polynucleotide strings in the genome of cells. Its structure is stabilized by hydrogen bonds between the bases in one strand and those of the other strand. Those hydrogen bonds can be broken with raising the temperature, but not the covalent bonds holding a chain together. The separation (or melting) of DNA strands is called DNA denaturation. DNA denaturation can be detected by ultraviolet absorption at 260 nm (single-stranded and double-stranded nucleic acids differ). Double-stranded can be achieved by cooling

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME







down slowly the denatured, single strand DNA. The re-form of native DNA is called annealing. When two strands come from different sources, than hybrid molecules can be formed. The later one is called hybridization - the artificial assembly of a double-stranded nucleic acid by complementary base pairing of two single strands.

RNA molecules are usually single-stranded. Its secondary structure is formed with folding of the molecule at complementary base pairing sites. The final three dimensional shape is critical for the biological function of the RNA.

Information stored in DNA is transferred to ribonucleic acid (RNA) during gene expression. The three steps of genetic information flow are replication, transcription, and translation. DNA double helix is duplicated during replication step. The genetic information is transferred from DNA to RNA in the transcription process. Translation refers the synthesis of a protein, using the information in messenger RNA (mRNA) in the cell.

Four different RNA classes can be found in the living organisms. (i) Messenger RNA (mRNA) is a single-stranded molecule, carrying the genetic information from DNA to the ribosome. (ii) Transfer RNAs (tRNAs) has important role in the translation by converting the genetic information from mRNA into proteins, while the (iii) ribosomal RNAs (rRNAs) are catalytic and structural components of the ribosome. In addition to these, cells contain a variety of (iv) small RNAs, which are ~21-26-nucleotide RNA molecules, and regulate either the production or the activity of proteins or other RNAs. Those are microRNA (miRNA) and small interfering RNA (siRNAs), which distinguished by their mode of biogenesis.

The three steps of genetic information flow are used in all cells and constitute the central dogma of molecular biology. Some viruses violate the central dogma using RNA as the genetic material and either replicate their RNA using RNA as template, or RNA genome is converted to a DNA version by a process called reverse transcription in retroviruses.

There is a linear correspondence between the base sequence of a gene and the amino acid sequence of a polypeptide. Three bases (called a codon) on an mRNA molecule encodes a single amino acid. This genetic code is translated into protein by the ribosomes (consist of proteins and rRNA), tRNA, and proteins known as translation factors.







### 3 DNA STRUCTURE AND REPLICATION, TRANSCRIPTION AND PROTEIN SYNTHESIS

The DNA is double-stranded molecule. Each chromosome consists of two strands of DNA in the cell. Each DNA strand contains several (hundreds of thousands to million) nucleotides. Nucleotides are linked by phosphodiester bonds. The strands are held together by hydrogen bonds that form between the purine and pyrimidine bases in one strand and those of the other strand. The two strands of DNA possess complementary in base sequence, which is ensured by specific base pairing: guanine (G) bonds with cytosine (C) and adenine (A) bonds with thymine (T). The two strands of the double-stranded DNA molecule are arranged in an antiparallel way: one strand runs 5' to 3', whereas the other strand runs 5' to 3'. The two strands form a double helix with wrapping around each other. There are two distinct grooves in the DNA helix: a major groove and a minor groove. Most DNA interacting protein binds in the major groove. B, A and Z double-helical structures are possible for DNA. The B-form or B-DNA is the double helix described by Watson and Crick. It has 10 base pairs per turn. The A-form is shorter and fatter, having 11 base pairs per turn and the major groove is narrower and deeper. It is often formed by double-stranded RNA or hybrids of one RNA plus one DNA strand. The sugar-phosphate backbone of Z-DNA double helix is a zigzag line, having 12 base pairs per turn, and it is left handed. Z-DNA is found in GC- or GT-rich regions, enzymes (occasionally) and regulatory proteins bind there preferentially.

The size of a DNA molecule is characterized by the number of nucleotide bases or base pairs per molecule. A DNA molecule with 1000 bases is 1 kilobase (kb) of DNA, while for the double-stranded DNA kilobase pairs (kbp) is used. The size of the DNA in the chromosome of *Escherichia coli* bacterium has about 4640 kbp, or 4.64 megabase pair (Mbp). 1 kbp of DNA is 0.34  $\mu\text{m}$  long with 100 helical turns, as one base pair takes up 0.34 nanometer (nm) in length, and each turn of the helix contains approximately 10 base pairs in the B-DNA. Therefore, The *E. coli* genome is thus  $4640 * 0.34 = 1.58 \text{ mm}$ , while a cell of *E. coli* is about 2  $\mu\text{m}$  long. Bacterial chromosome is structured with supercoiling, when the double-stranded DNA is further twisted. Supercoiling puts the DNA molecule under torsion (like the added tension to a rubber band that occurs when it is twisted). DNA can be supercoiled in either a positive or a negative manner (when the DNA is twisted about its axis in the opposite sense from the right-handed double helix). Negatively supercoiled DNA is the form predominantly found in nature.

Supercoils are inserted or removed by topoisomerase enzymes. There are two major classes of topoisomerases: class I and II. Class I topoisomerases make a single-stranded break in the double helix, allowing the rotation of one strand around the other. One rotation adds or removes one single supercoil. Class II topoisomerases (e.g. gyrase) make double-stranded breaks. Then double helix is passed through the break. Afterwards the break is resealed, and two supercoils added or removed. Energy provided by ATP is necessary for inserting supercoils, but not for releasing. DNA molecule can be alternately supercoiled and relaxed. Supercoiling is necessary for packing the DNA into the cell. Relaxation of DNA is necessary for its replication and transcription. There are special antibiotics inhibiting DNA gyrase (e.g. nalidixic acid, ciprofloxacin, and novobiocin).





DNA replication is necessary for cell division to reproduce new organisms in case of unicellular microorganisms, or to produce new cells in multicellular organism. DNA replication must be accurate to provide the genetic identity of daughter cells to the mother cell. DNA replication is semiconservative, meaning that one new strand and one parental (template) strand build each of the two resulting double helices. New DNA strands are building from a deoxynucleoside 5'-triphosphate precursor molecules. The two terminal phosphates are removed from the precursor deoxynucleoside 5'-triphosphate, then the 5'-phosphate of the incoming nucleotide being attached to the 3'-hydroxyl of the previously added nucleotide. This means that DNA replication always proceeds from the from 5' end to 3' end.

DNA polymerase enzymes catalyse the addition of deoxynucleotides. Several DNA polymerases exist, in cells. There are DNA polymerases I, II, III, IV, and V DNA polymerases in *Escherichia coli*. DNA polymerase III (Pol III) is the primary enzyme for replicating chromosomal DNA, and DNA polymerase I (Pol I) also participate in chromosomal replication. The other DNA polymerases help repair damaged DNA. All known DNA polymerases can only add a nucleotide onto a preexisting 3'-OH group but no known DNA polymerase can initiate a new chain. To start a new chain, an oligonucleotid primer (usually a short stretch of RNA) is required, to which DNA polymerase can attach the first nucleotide. Primase enzyme, synthesizes a short stretch of RNA of around 11–12 nucleotides that is complementary in base pairing to the template DNA, when the double helix is opened at the beginning of replication. DNA polymerase can add the first deoxyribonucleotide at the growing, 3'-OH end of this RNA primer.

Replication occurs at replication fork, where the DNA is unwound. The DNA helicase enzyme unwinds the double helix, using energy from ATP, and exposes a short single-stranded region. Helicase moves along the DNA and separates the strands just in advance of the replication fork. The single-stranded DNA region is stabilized by single strand binding protein. Those are covering ss DNA, and preventing from re-forming double helix. To counteract positive supercoils generated by helicase with unwinding of the double helix, DNA gyrase inserts negative supercoils to cancel out the positive supercoiling ahead of the replication fork.

Bacterial chromosome has only one single replication origo (*oriC*) where DNA synthesis is initiated. This consists of a specific DNA sequence of about 250 bases that is recognized by initiation proteins, which binds to this region and opens up the double helix. Initiation of DNA replication then begins on the two single strands. As replication proceeds, the replication fork appears to move along the DNA in two directions. On the leading strand growing occurs continuously from the 5'-PO<sub>4</sub><sup>2-</sup> to the 3'-OH, because there is always a free 3'-OH at the replication fork. But on the opposite strand, called the lagging strand, DNA synthesis occurs discontinuously because there is no 3'-OH, but a 5'-PO<sub>4</sub><sup>2-</sup> at the replication fork to which a new nucleotide can attach. DNA synthesis is discontinuous at the lagging strand, as RNA primers synthesized multiple times, providing free 3'-OH. As a result, the lagging strand is made in short segments, called Okazaki fragments, after their discoverer, Reiji Okazaki. These fragments are joined together with the help of Pol I and DNA ligase enzymes to give a continuous strand of DNA. Pol I enzyme removes the RNA primer with its 5' - 3' exonuclease activity, and replace with DNA synthesizing complement DNA strand. Finally DNA ligase seals nicks in DNAs with binding the adjacent 5'-PO<sub>4</sub><sup>2-</sup> and 3'-OH of the strands. The opposite side of the circular bacterial chromosome from the origin is called the terminus of replication. Here the two replication forks collide as the new circles of DNA are completed, and the process of DNA replication is finished.

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



Transcription is the synthesis of ribonucleic acid (RNA) using DNA as a template. Transcription is carried out by the enzyme RNA polymerase, catalyzing the formation of phosphodiester bonds between ribonucleotides, using ribonucleoside triphosphate precursors, and DNA as a template. Ribonucleoside triphosphates are added to the 3'-OH end of the ribose of the previous nucleotide during the elongation of the RNA chain, and energy is provided by the two energy-rich phosphate bonds of the incoming ribonucleoside triphosphates. RNA polymerase can initiate new strands of nucleotides on its own (primer is not necessary). Only one of the two strands is transcribed for any given gene from the double-stranded DNA template during transcription.

RNA polymerase first recognizes the initiation site on the DNA, called promoter. Then it binds to the promoter and transcription can proceed to initiate RNA synthesis. During the elongation, DNA double helix is opened at the promoter (by RNA polymerase), and a transcription bubble is formed. While polymerase moves, it unwinds the DNA in short segments, and copy into the RNA complement. Transcription stops at specific sites called transcription terminators. Newly made RNA dissociates from the DNA, and the opened DNA closes back into the original double helix form.

Transcription copies much smaller units of DNA (often only a single gene), than DNA replication. This system allows the cell to transcribe different genes at different frequencies, depending on the needs of the cell for different proteins. Genetic information on chromosomes is organized into transcription units (segments of DNA that are transcribed into a single RNA molecule). Transcription units include either only one gene, or two or more genes. These genes are said to be cotranscribed, yielding a single RNA molecule.

In prokaryotes, genes encoding related enzymes are often clustered together in operon, from where genes are transcribed together to give a single, long mRNA. An mRNA encoding such a group of cotranscribed genes is called a polycistronic mRNA. Then, during the translation of the polycistronic mRNA, several polypeptides are synthesized, by the same ribosome. Operon allows the coordinated expression of genes for the same biochemical pathway or genes needed under the same conditions (e.g. *gal*, *trp* and *his* gene cluster in *E. coli*).





## 4 EUKARYOTIC GENETICS AND MOLECULAR BIOLOGY

---

The eukaryotes have linear chromosomes in the nucleus. The presence of nucleus results the physical separation of transcription and translation, and the need to disassemble the nucleus during cell division. The replication of eukaryotic linear chromosomes and the processing of mRNA inside the nucleus are largely.

Protein-encoding genes in eukaryotes are often containing introns. Introns are noncoding regions, which are splitting coding regions (exons) into two or more parts. In prokaryotes only a few genes contain introns. Eukaryotic microorganisms have fewer introns than do higher eukaryotes. For example, about 70% of DNA in the yeast *Saccharomyces cerevisiae* encodes protein, while in the human genome only about 3% of the total DNA encodes protein. The non-coding DNA sequences in eukaryotes are present as introns or repetitive sequences, which may be repeated hundreds or thousands of times. Additionally, eukaryotes often have multiple copies of certain genes, like tRNAs, rRNAs and histon encoding ones.

Both introns and exons are transcribed into the primary RNA (pre-mRNA) transcript. Introns are removed and the exons are spliced together during maturation forming a functional mRNA. Only the matured mRNA exits the nucleus.

Eukaryote has multiple linear chromosomes inside the nucleus; while prokaryotic chromosome is circular, and usually a single one can be found in its cytoplasm. Because eukaryotic chromosomes are in the nucleus and the ribosomes are in the cytoplasm, transcription and translation are spatially separated processes.

In eukaryotes, large amounts of protein are bound to the DNA. The linear DNA molecules of the eukaryotic chromosomes are wound around histone proteins to form structures called nucleosomes. Each nucleosome contains approximately 200 base pairs of DNA plus nine histone proteins: two from each of the four core histones (H2A, H2B, H3, and H4) and one linker histone, H1. Eukaryotic core histones are the most highly conserved during evolution of all known proteins. For example, only two amino acids out of 102 are different between histone H4 of cows and peas.

The complex of DNA and histone proteins is called chromatin. It can be further compacted by folding and looping forming a dense structure, called heterochromatin. The compacted chromosomes are visualized in eukaryotic cells during cell division.

Eukaryotic cells can exist in two forms regarding the number of their chromosomes: haploid or diploid. Haploid cells have one copies of each chromosome whereas diploid cells have two from each. Many single-celled eukaryotes, such as the brewer's yeast (*Saccharomyces cerevisiae*), exist mainly as haploid cell (containing 16 chromosomes). Occasionally, two haploid yeast cells will fuse (mate) to yield a diploid cell (with 32 chromosomes). On contrary, the diploid phase is present in the somatic cells of plants and animals, and the haploid phase occurs only transiently, in the gametes.

Haploid and diploid somatic cells are multiplying following DNA replication in the process called mitosis. During mitosis, the

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



chromosomes condense, divide, and are separated into two sets, one for each daughter cell.

The other process of eukaryotic cell division is meiosis, when the diploid cells convert to the haploid stage. Meiosis consists of two cell divisions: homologous chromosomes segregate into separate cells, changing the genetic state from diploid to haploid in the first meiotic division. It is followed by a second meiotic division, where the two haploid cells divide to form finally a total of four haploid gametes.

During transcription, first initial RNA produced as the primary transcript in eukaryotes. However, RNA must be matured (known as RNA processing), to carry out their role in the cell. The three processes of RNA maturation is (i) capping, (ii), splicing, and (iii) getting poly A tail. Capping is the addition of a methylated guanine nucleotide at the 5'-phosphate end of the mRNA, and occurs before transcription is complete. The cap nucleotide is added in reverse orientation relative to the rest of the mRNA molecule. The guanosine cap promotes the formation of the initiation complex between the mRNA and the ribosome through specific cap-binding proteins during translation.

Polyadenylation of mRNA is the process, when 100–200 adenylate residues are added to the 3' end of the primary transcript as the poly(A) tail. The tail recognition sequence (AAUAAA) is located beyond the stop codon, close to the 3' end of the primary transcript. The poly(A) tail is required for translation; it indicates to the translation machinery that the RNA is mRNA and that it is ready for translation. Moreover the poly(A) tail must be removed before the mRNA can be degraded, so it stabilizes mRNA. The mRNA is transported into the cytoplasm for translation, only when all three steps of maturation (capping, splicing and polyadenylation) are complete.

Many genes in eukaryotes contain the introns. Introns are removed during splicing and exons are joined to form a contiguous protein-coding sequence in the mature mRNA. RNA splicing is in the nucleus, and it is done by a large macromolecular complex, called the spliceosome. The spliceosome is about the size of a ribosome, and contains four large RNA–protein complexes. Each snRNPs contains noncoding small nuclear RNA (snRNA) that recognize conserved sequences (at the splice junctions) on the primary transcript by base pairing. The proteins of the spliceosome cut out the intron, and join the flanking exons together. The intron is removed as a lariat structure that is then degraded. There are self-splicing introns as well, called ribozymes. They belong to RNAs having enzymatic activity and, participate in several cellular reactions, like splicing. Self-splicing introns are introns that fold up to generate three-dimensional structures with ribozyme activity. This enzymatic activity allows them to excise themselves from an RNA molecule while joining adjacent exons together. Most self-splicing introns are found in the genes of mitochondria and chloroplasts.

Eukaryotes have multiple RNA polymerases, unlike Bacteria and Archaea, which have just one. The eukaryotic nucleus contains RNA polymerases I, II, and III, which transcribe different categories of nuclear genes. Moreover the RNA polymerases of the eukaryotic nucleus have ten or more subunits, while Bacteria contain a relatively simple (five-polypeptide) RNA polymerase. The structure of the mitochondrial and chloroplasts' RNA polymerase are much more similar to the bacterial one. Protein-encoding genes are transcribed by eukaryotic RNA polymerase II. RNA polymerase I transcribes the 18S and 28S rRNA, and RNA polymerase III transcribes genes for tRNA, 5S rRNA, and other small RNA molecules. Each RNA polymerase recognizes a distinct class of promoters. There are two key elements of the eukaryotic promoters, a TATA box, and an initiator element. RNA polymerase II is binding to a promoter on the DNA contains a TATA

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



box. Initiator element (INIT) is very near to the transcription start site. Eukaryotic RNA polymerases require transcription factors to recognize specific promoters. General transcription factors are needed for the functioning of all promoters recognized by a particular RNA polymerase, while specific transcription factors are needed for expression of certain genes under specific circumstances.

Protein synthesis by eukaryotic ribosomes is generally more complex than in Bacteria. The cytoplasmic ribosomes of eukaryotic cells (80S ribosomes) are larger than bacterial ribosomes (70S ribosomes) and contain more rRNA (5S, 5.8S, 28S in the large subunit and 18S in the small subunit) and protein molecules, than the bacterial one (5S, 23S in the large subunit and 16S in the small subunit). Eukaryotes also have more initiation factors and a more complex initiation procedure. However eukaryotic cells also contain bacterial-type ribosomes in their mitochondria and chloroplasts. In eukaryotes, mRNA carries only a single gene that is translated into a single protein (monocistronic), while in Bacteria mRNA is polycistronic, and may be translated to give several proteins.

Initiation of protein synthesis differs significantly in eucariotes. Its special initiator tRNA recognizes the start codon and inserts methionine, the first amino acid. Bacteria use N-formylmethionine as the first amino acid of all proteins, and different initiator. Eukaryotic mRNA is recognized by its cap, as a specific protein, cap-binding protein, binds both the mRNA cap and the ribosome. However bacterial eukaryotic mRNA has ribosome-binding site (Shine–Dalgarno sequence). The first AUG to be found on the mRNA is normally used as the start codon in eukaryotes. The order of assembly of the ribosomal initiation complex is different, as in eukaryotes the initiator Met-tRNA binds to the small ribosomal subunit before the mRNA, but the opposite happens in Bacteria. Eukaryotic translation is sensitive to difteria toxin (produced by *Corynebacterium diphtheria*), while it does not inhibit the bacterial process.







## 5 PRINCIPLES OF BACTERIAL GENETICS AND PROKARYOTIC GENOMES

Several bacterial genomes have been completely sequenced and analysed, thus revealing the number and location of the genes they possess. The MG1655 of *Escherichia coli*, whose chromosome was originally sequenced, is a derivative of the traditional *E. coli* K-12 strain, used for genetics. Its chromosome contains 4,639,221 bp, with 4288 possible protein-encoding genes that account for about 88% of the genome, while tRNAs and rRNAs encoding genes are approximately 1%. Regulatory sequences (promoters, operators, origin and terminus of DNA replication, etc.) comprise around 10%, and only the remaining 0.5% of genome consists of noncoding, repetitive sequences.

Genes that encode the enzymes of a single biochemical pathway in *E. coli* are often clustered (e.g. gal, trp, and his gene cluster). Each of these gene clusters constitutes an operon that is transcribed as a single, polycistronic mRNA carrying multiple coding sequences. However over 70% of the 2584 predicted or known transcriptional units in *E. coli* are not clustered, and contain only a single gene. Only about 6% of the operons have four or more genes.

The direction of transcription can be either clockwise or counter clockwise on the prokaryotic circular chromosome. This means that some coding sequences are on one strand of the chromosome whereas others are on the opposite strand. There are about equal numbers of genes on both strands in *E. coli*, but highly expressed genes (like all rRNA operons and majority of tRNA genes) in *E. coli* are transcribed in the same direction that the DNA replication fork moves. The two replication forks start at *oriC*, and move in opposite directions around the circular chromosome toward the terminus.

Approximately 4225 different proteins may be encoded by the *E. coli* chromosome, indicated by the sequence analyses of its genome, but 30% of these hypothetical proteins has still unknown function. The largest gene in *E. coli* encodes a protein of 2383 amino acids that is still uncharacterized. Mutant isolation, biochemical and physiological analyses can help finding the function of these hypothetical proteins.

Gene duplication plays a major role in evolution, like in the evolving of gene families (groups of genes with related sequences and encoding products with related functions). Gene families are common, both within a species and across broad taxonomic lines. Computer analyses have shown that several protein-encoding genes arose by gene duplication in *E. coli*. Groups of genes with related sequences encoding products with related functions (called gene families) also can be found in the genome of *E. coli*. For example, the large family of membrane transport proteins contains 70 genes.

Several inserted genetic elements can be found in the *E. coli* chromosome. There are multiple copies of several different insertion sequences (IS elements), and defective integrated viruses (vary from nearly complete virus genomes to small fragments).

*E. coli* obtained nearly 20% of its genome by horizontal (lateral) gene transfer from other organisms. Horizontal gene transfer may cause large-scale changes in a genome. For example, virulence genes located on pathogenicity islands can be acquired by horizontal transfer. Many genes acquired in this way provide no selective advantage and so are lost by deletion. This keeps

TÁMOP-4.1.1.C-12/1/KONV-2012-0014

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

SZÉCHENYI 2020

Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE





the chromosome of a given species at roughly the same size over time. Genome size is therefore a species-specific trait, e.g. 4.5–5.5 Mbp for *E. coli*.

Prokaryotic genome may contain plasmids, in addition to the chromosome. Plasmids are replicating independently of the host chromosome. However, they do rely on chromosomally encoded enzymes for their replication. Plasmids exist inside cells as free, typically circular DNA, and do not have an extracellular form (like viruses). Plasmids carry only nonessential (but often very helpful) genes, while essential genes reside on chromosomes. Thousands of different plasmids are known (over 300 in *E. coli* alone). Plasmids have been widely exploited in genetic engineering. Several artificial plasmids have been created to incorporate genes from a wide variety of sources, allowing their transfer across any species barrier.

Most plasmids are circular, but many linear plasmids are also known, and almost all consist of double-stranded DNA. The size of the naturally occurring plasmids varies from 1 kbp to more than 1 Mbp. A typical plasmid is circular, double stranded DNA, and its size is less than 5% of the host chromosome. Some bacteria may contain several different types of plasmids (e.g. the Lyme disease pathogen *Borrelia burgdorferi* (contains 17 different circular and linear plasmids).

Plasmids are using normal cellular enzymes for replication. Only genes controlling the initiation of replication and partitioning replicated plasmids between daughter cells are carried by the plasmid itself. Plasmids are present in cells in different numbers (copy number). Copy number is controlled by genes on the plasmid and by interactions between the host and the plasmid. Copy number may vary between from 1–3 copies to over 100 copies.

Several replication methods have been described for plasmids. Most plasmids in gram-negative Bacteria replicate in a manner similar to the bacterial chromosome: replication starts from one origo and continue in bidirectional around the circle (a theta intermediate). Some small plasmids have unidirectional replication, with a single replication fork. Most plasmids of gram-positive Bacteria, and a few gram-negative Bacteria, replicate by a rolling circle mechanism similar to some bacteriophage, via a single stranded intermediate. Most linear plasmids however replicate by using a protein bound to the 5' end of each strand to prime DNA synthesis. Episomes are plasmids, which can integrate into the chromosome. Their replication is controlled by the chromosome.

Bacterial cells may contain multiple different plasmids. But plasmids, which are genetically closely related, cannot be maintained in the same cell (incompatible). When a plasmid is transferred into a cell that already carries another related and incompatible plasmid, one of them will be lost during subsequent cell replication. Therefore, although a bacterial cell may contain different kinds of plasmids, each is genetically distinct.

Plasmids can be eliminated from host cells with inhibiting plasmid replication without parallel inhibition of chromosome replication. It is called curing, which may occur spontaneously, but is can be greatly increased by treatments with certain chemicals (e.g. acridine dyes).

Few bacterial species cells can take up free DNA from the environment, so plasmids released by previous host cell may be taken up by a new host. However the main mechanism of plasmid transfer is conjugation. This function involves cell-to-cell contact, and is encoded by some

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



conjugative plasmid itself. Transfer by conjugation is controlled by *tra* (for transfer) genes on the plasmid called. These genes encode proteins that function in DNA transfer, replication or in mating pair formation. Most conjugative plasmids can only move between closely related species of bacteria. However, some conjugative plasmids from *Pseudomonas* are transferable to a wide variety of other gram-negative Bacteria. Moreover conjugative plasmids have been shown to transfer between gram-negative and gram-positive Bacteria, between Bacteria and plant cells, and between Bacteria and fungi, and the genome of the new host may be recombine with them. Plasmids are a major mechanism for conferring special properties on bacteria and for mobilizing these properties by horizontal gene flow.

Plasmids may carry genes that influence host cell physiology, like conjugation, resistance, virulence, antibiotic production, or ability of special metabolic function. For example, certain plasmid encoded proteins provide interaction between *Rhizobium* and plants to form nitrogen-fixing root nodules. The ability to degrade toxic pollutants (e.g. octane, camphor, or naphthalene) by *Pseudomonas* is also coded in by plasmids.

Resistance plasmids (usually called R plasmids) are one of the most widespread and well-studied groups. They confer resistance to antibiotics and various other growth inhibitors. In general, resistance genes encode proteins that either inactivate the antibiotic or protect the cell by some other mechanism. Multiple resistance of a bacterial cell is resulted either because the presence of a single R plasmid possessing several antibiotic resistance genes, or one cell may contain several R plasmids. Plasmid R100, for example, is a 94.3-kbp plasmid, that carries genes encoding resistance to sulfonamides, streptomycin, spectinomycin, fusidic acid, chloramphenicol, and tetracycline, moreover several genes conferring resistance to mercury. Soon after antibiotic resistant strains were isolated, it was shown that they could transfer resistance to sensitive strains via cell-to-cell contact, and the conjugative R plasmids permitted their rapid spread through cell populations. For example R100 plasmid can be transferred between enteric bacteria (*Escherichia*, *Klebsiella*, *Proteus*, *Salmonella*, and *Shigella*).

The two major characteristics of the virulence (disease-causing ability) of pathogens that are often plasmid encoded in the pathogenic microorganisms: (1) the ability of the pathogen to attach to and colonize specific host tissue and (2) the production of toxins, enzymes, and other molecules that cause damage to the host. In the enteropathogenic strains of *E. coli* colonization factor antigen, encoded by a plasmid. This protein confers on bacterial cells the ability to attach to epithelial cells of the intestine. Moreover at least two toxins are encoded by plasmids in enteropathogenic *E. coli*: the hemolysin (lyses red blood cells), and the enterotoxin (responsible for diarrhea).

Many bacteria produce bacteriocins, proteins that inhibit or kill closely related species or even different strains of the same species. The genes encoding bacteriocins and other proteins needed for processing and transporting them, moreover the ones conferring immunity on the producing organism are usually also carried on plasmids colicins (bacteriocins produced by *E. coli*) are encoded on Col plasmids. Col plasmids can be either conjugative or nonconjugative. Colicins kill cells by disrupting some critical cell function: forming channels in the cell membrane, degrade DNA or RNA. Nisin A is a bacteriocin produced by lactic acid bacteria, which strongly inhibits the growth of a wide range of gram-positive bacteria have commercial value, as it is used as a preservative in the food industry.





## 6-7 REGULATION OF GENE EXPRESSION I-II

The cells are basic organizational and functional units of life. Their identity and functions are determined by their constituent protein molecules. Proteins are formed based on the genetic information encoded in DNA. This is called gene expression. One gene encodes one protein. If the number of the genes is known in the genome of a species, and we know the number of the proteins in a cell of the species, it is clear that only a small part of the genes are active. Consequently the cells regulate the activity of the genes by decreasing or increasing their transcription. It is essential that the gene expression can be regulated in such a way that the corresponding gene products are obtained in the cell at the right time.

Small molecules often participate in the control of the transcription, however those rarely exert their effects directly. Instead, those affect the binding of specific proteins to DNA. These proteins are called regulatory protein or transcription factors which are small protein molecules, the majority of those are sequence-specific. The regulatory proteins are able to form complex with the operator region which is located upstream of the regulated gene. Structurally the regulatory proteins are homodimers which means that they consist of two identical subunits. Each subunit has a domain, which is responsible for the DNA binding. These domains recognise the complementary sequence thus these domains bind to opposite strands. The DNA binding domains of the transcription factors often fall into structural categories. We distinguish three types of structures: helix-turn-helix, zinc finger, leucine zipper.

Prokaryotic gene expression has the following characteristics:

- translation and transcription occur in a common area, closely follow each other
- regulation of transcription is done mostly by repressor proteins
- the promoter and the RNA polymerase is usually enough to start the transcription
- several protein encoding open reading frame (ORF) per transcription units
- polycistronic mRNA

The operon model originally it was considered a control mode of prokaryotes, however this kind of control have also been identified in some eukaryotic organisms.

Operon is a functionally integrated genetic unit in DNA, which contains a promoter site, an operator region, structural genes and a terminator region. The transcription regulatory proteins (activators or repressors) bind to the operator. The RNA polymerase binds to the promoter. Depending whether the regulatory protein is an activator or repressor, the binding of the RNA polymerase to the promoter assisted or inhibited. If more than one operon is under the control of a single regulatory protein, these operons are collectively called regulon.

The structural genes are located one after the other and are regulated together and usually encode enzymes of a biochemical pathway. The RNA polymerase synthesizes one single, long mRNA, which contains all of the structural genes until the terminator region. This is called polycistronic mRNA.

There are two important elements of the prokaryotic promoter: the Pribnow-box also called -10 element or TATA box (TATAAT consensus sequence), which is located at 10 bases to the 5' direction from the first transcribed base pair (position +1). The presence of this element is essential for the transcription. The second element is another consensus sequence

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME





TTGACA which is also called -35 sequence, because it is located at 35 bases to the upstream direction from the first transcribed base pair. The presence of this element in the promoter is not essential, however it contributes significantly to the transcription of the gene. Furthermore, many prokaryotic promoters contain additional nonessential sequences ("UP" elements) to the 5' direction from the -35 element.

If there is no specific mechanism to regulate gene transcription, the transcriptional frequency of a specific gene is determined by the strength of the promoter.

In case of a negative regulator the transcription regulatory protein is a repressor, which inhibits the synthesis of mRNA. Frequently, the enzymes that catalyse the formation of a specific product are not produced, if the product is present in large quantities in the medium. This is called enzyme repression. Alternatively, the enzyme can only be synthesized when its substrate is present in the medium. This is called enzyme induction. The enzyme repression typically influences biosynthetic (anabolic) enzymes while the enzyme induction affects the breakdown (catabolic) enzymes. The substance which induces the enzyme synthesis is the inducer, which represses the enzyme synthesis is called corepressor. These molecules together form the group of effectors. It is important to know that not all of the effectors are substrate or final product, it may also be structural analogue of these (e.g.: IPTG see below).

The operon is inducible, if the repressor protein binds to the operator, until the inducer molecule appears. As the inducer molecule binds to the repressor inactivates it.

Beta-galactosidase enzyme of *Escherichia coli* which splits lactose into glucose and galactose are not always present in cells. It is produced only in the presence of lactose, so the appearance of the enzyme is triggered by the substrate. In 1950's Francois Jacob and Jacques Monod studied the *E. coli* lactose metabolism and were looking for the answer to the question of how the cell knows when to produce the appropriate enzymes. They were awarded the Nobel Prize for their work.

The following enzymes are necessary for lactose utilization:

- Beta-galactosidase (*lacZ* gene product) - cleaves lactose molecule
- permease (*lacY* gene) - promotes the transport of lactose into the cell
- transacetylase (*lacA* gene) - the role is still unknown

The induction of galactoside activity can also be made by artificial galactoside derivatives. The best inducer is the IPTG (isopropyl beta-D-thio-galactopyranoside), which is not substrate of the enzyme. IPTG has an advantage that it does not lose during the induction progresses as the original inducer, the lactose. Removal of the inducer stops the enzyme synthesis. Thus the cells are able to turn on or turn off their genes due to changing environmental effects.

For the study of the enzyme activity two derivatives are used as substrates. Both are chromogenic substrates. It means that the specific enzyme and the substrate form coloured enzyme-substrate complex together. The colour intensity is proportional to the amount of the enzyme, thus it is easy to follow the activity. ONPG (o-nitrophenyl beta-D-galactopyranoside) is used in accurate measurements of enzyme activity and the amount of the resulting yellow product can be measured with a photometer. The other chromogenic substrate, the X-gal (5-bromo-





4-chloro-3-indolyl-beta-D-galactopyranoside) is added to the culture medium and the growing bacterial colonies containing beta-galactosidase turn blue. The presence or absence of the enzyme is easily detectable by blue-white selection. The strains without beta-galactosidase activity and the beta-galactosidase overproducing strains are easily distinguishable: the *lacZ*- mutant forms white colonies in induced state and the overproducing strains provide darker blue colonies than the wild-type.

The *lacI* gene encodes the repressor protein (LacI), which inhibits the *lacZYA* genes; beta-galactosidase production is beneficiary only in the presence of the inducer. The LacI has two binding sites which recognize the inducer and operator. It is an allosteric protein bound to the inducer causes its conformational change that removes the repressor-operator binding.

In case of repressible operon, the repressor protein is present in the cell, but does not bind to the promoter. If a co-repressor molecule binds to the repressor, the repressor undergoes conformational change which enables it to bind to the operator. Such is the *trp* operon which performs tryptophan biosynthesis: the end product inhibits the activity of enzymes involved in the biosynthesis (feedback mechanism). In the presence of the tryptophan the TrpR repressor binds to the operator site, so tryptophan is a co-repressor molecule. Tryptophan regulates the TrpED complex through a negative feedback. In this case, the accumulation of the final product "turns off" the biosynthesis.

If glucose and lactose are simultaneously present in the medium, there is strong *lac* expression. This phenomenon is called catabolite repression (global repression). The glucose is a better carbon and energy source, so bacteria use it previously, thus the enzymes for lactose utilization are not required. This also happens with other carbon sources. If the cell is "starving", few ATP and many cyclic AMP (cAMP) are formed. The cAMP and the catabolite activator protein (CAP) form a cAMP-CAP complex. This complex activates other sugar utilization genes such as the *lac* operon as well. If the glucose is enough, the cAMP:ATP ratio is shifted in favor of the ATP, thus there are not enough cAMP-CAP complexes for the activation of other sugar utilization operons (e.g.: *lac*, *ara*). This control is an example that the same protein is able to operate as activator and repressor as well.

In contrast to the negative control the regulatory protein is an activator that promotes binding of the RNA polymerase and the DNA. Normally the RNA polymerases bind weakly to DNA. The activator proteins help RNA polymerases to recognize the promoter and to start transcription or modify the structure of DNA. These proteins can also modify the structure of RNA polymerase in order to be able to connect to the promoter.

Normally the activator protein does not bind to the operator in the positive inducible operon. Before the binding it must connect to an inducer molecule. Connecting to the inducer leads to the binding of the regulatory protein to the DNA.

In case of positive repressible operon, normally the activator molecule binds to the operon and promotes the transcription. Appearance of an inhibitor molecule and its connection with the activator protein prevents the additional binding of activator to the operator.

The major characteristics of transcriptional regulation in eukaryotes:

- mechanism of regulation is similar as in prokaryotes
- transcription and translation take place in different compartments, separated from each other

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



## DEBRECENI EGYETEM

- in the regulation of transcription mostly activator proteins are involved: the promoter and the RNA polymerase (without activator protein) are not sufficient to start the transcription
- three types of RNA polymerase: pol-I, pol-II, pol-III
- one protein encoding open reading frame (ORF) per transcription units
- monocistronic mRNA
- mRNA undergoes splice before leaves the nucleus (an average gene is considerably longer than the matured mRNA: exons, introns, splicing)

The majority of eukaryotic genes have promoters, in which the TATA box is present in the same way, as in the prokaryotic genes. Transcription factors recognize the characteristic DNA base pair sequences (cis-regulatory targets) and link to them. Approximately 250 types of transcription factor binding sequences are known and up to a dozen may also be in the promoter of a gene. Various transcription factors that bind with different strength to the DNA are often competing for the same sequence. The fact that a gene is turned on or turned off and how often is transcribed by RNA polymerase depends on the kind and the number of transcription factors in the cell and their binding ability (number of binding site and bond strength).

The bacterial cells are exposed to changing environmental conditions. For their survival it is essential to preserve the stability of their internal environment (homeostasis). Moreover, energy saving is very important, which means that the synthesis of unnecessary, wasteful proteins have to be prevented. Their genes are often regulated in response to external signals. If the signals coming from the environment are small molecules, they are able to function as effectors in cells. However, the cells also receive other signals, such as pH, oxygenation and temperature. These signals are detected and transmitted by the sensors. This process is called signal transduction.

Most of the signalling systems consist of two parts. One of these includes proteins with sensor and kinase activity (sensor kinases), and cytoplasmic response regulator proteins (regulators) belong to the other part. The sensor kinases detect the signals from the environment and phosphorylate themselves (autophosphorylation) at a specific histidine residue (sensor kinases belong to the group of histidine kinases). Thereafter the phosphate is transmitted to a response regulator protein, which is generally a DNA-binding protein that regulates the transcription (positive or negative manner). A phosphatase enzyme (capable of removing phosphate) is also present in the control cycle. This phosphatase continuously removes the phosphate from the response regulator proteins at a constant speed; thereby it is possible to eliminate the response (feedback loop). The phosphatase activity is typically slower than the phosphorylation. There are proteins that are able to simultaneously detect the signal and to carry out the control (such as the LacI).

The two-component regulatory systems are present in microbial eukaryotes such as *Saccharomyces cerevisiae*, and plants as well. However, the most eukaryotic signalling pathways are based on the phosphorylation of serine, threonine and tyrosine residues of the proteins, which is not related to bacterial two component regulatory systems. In prokaryotes, a modified two-component system is used to regulate the chemotaxis. Using this regulatory mechanism they are able to control the rotation of flagellum depending on the presence of attractive or repellent materials.

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**





Most bacteria is capable of monitoring the cell density of the population, which is accomplished by production and detection of low molecular weight signal molecules (autoinducers). This phenomenon is called quorum sensing, which means cell density-dependent coordinated gene expression. Their goal is the efficiency. For example, if a single cell begins toxin production, it only wastes resources. In contrast, the coordinated toxin production of a sufficiently large population may cause disease. All cells of the population emit the same autoinducer molecule that diffuses freely into cells. The autoinducer is only able to achieve high concentration in an individual cell, if many cells are present in the environment. When the concentration reaches the critical level, the gene expression pattern of the total population changes at the same time by turning off and turning on specific genes. The phenomenon is most prevalent among gram negative microbes, but it can be found in some gram-positive microbes, as well.

If the cells are starving, synthesis of rRNA and tRNA is inhibited. In this case, two unusual nucleotide derivatives appear. These originate from GDP and GTP, and wear 3' pyrophosphate moiety: guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp); together they are named (p) ppGpp (alarmones in *Escherichia coli*). The RelA protein (linked to the 50S ribosomal subunit) is responsible for their formation. RelA adds phosphate group to the GDP and GTP from ATP in the case of amino acid limitation. The ppGpps bind RNA polymerases and prevent the transcription.

The RNA alone, without regulatory proteins is capable of controlling the gene expression also at transcriptional and at translational level. Those RNAs that are not translated collectively are called non-coding RNA are (ncRNA). ncRNAs include tRNAs, rRNAs and small RNAs (sRNA). The different types of ncRNAs are able to exert their effects at several levels of the expression of genetic information. Their attack points can be the changing of the chromatin structure, the DNA imprinting, transcription and translation. The small RNAs are about 40 to 400 nucleotides in size. They regulate the prokaryotic and eukaryotic gene expression, as well. In the regulation of gene expression the small RNAs linked to the complementary mRNA are able to induce degradation of mRNAs, or inhibit their translation.







## 8 BACTERIAL GENE MANIPULATION

Each of the microorganisms used in the fermentation industry originates from a wild strain. The wild strains are usually isolated from nature, usually from the soil, but they may come from natural waters, from mud, from living and dead plants and from animal materials. The selection may be done at places, where the circumstances which are important for us (product, properties, transformation process) are present. For example, chemicals that deplete contaminating microbes are more likely to be isolated from chemically contaminated soils, waters, waste water treatment plants etc. or insect pathogens can be isolated from alive and dead insects. In another case, it may be expedient to isolate microbes which belong to the same race or which are very similar in racial kinship to microbes about which we know to be suitable for a particular task. Of course it is possible that we are looking for a microbe which is suitable for a completely new industrial task. The screening aims are the finding of the desired strain, metabolite, enzyme, or biotransformation pathway. Using several test methods we test the microorganisms involved in the experiment. The number of such microorganisms may range from a few dozens to several tens of thousands. In simpler cases, the filtering can be done in Petri dishes, by the investigation of strains grown in medium. However, most of the screening work is a serial test done in small tubes or on microtiter plates / sheets. In this case, it is tested the biological activity of the fermentation broths or metabolites separated from them.

It is rarely enough only to choose the best sub-strains, since the wild type strains never synthesize in large quantities of the compound in question (they are not able to overproduce), so it is necessary to develop them specifically. One way of doing this is the *in vivo* genetic modification, which is based on the use of classical mutation techniques and the selection of the resulting mutants and variants. Variations in small extent of course are created and selected also among natural circumstances. This is called phylogenesis. However, where the goal is the isolation of a microorganism capable of a given task, we need to increase the frequency by using mutagenic agents (methylnitro-nitrosoguanidine, ethyl methanesulfonate, and nitrogen mustard), X-rays or ultraviolet irradiation. For industrial development of strains the most commonly used is the UV radiation. The resulting mutants consequently are filtered again, their breeding circumstances are optimized, and the selected strains are developed further.

The accidental *in vivo* modification or manipulation of the phenotype of micro-organisms (creating random mutations) was used long before the genetic background of phenotypic characteristics has been recognized. However, knowledge of the genes led to the spreading of *in vitro* manipulation techniques (molecular cloning), changing certain genes, or inserting foreign genes into the genome in the hope that the change results in a useful phenotype. Although nowadays the main focus is on the *in vitro* methods, it is worth noting that sometimes the classical methods are actually easier and faster, especially if the genome of the organism is less known.

Classical mutagenic strategies use different methods. Alkylating agents, such as ethyl methanesulfonate (EMS), 1-methyl-3-nitro-1-nitrosoguanidine (MNNG) and mustard gas are extremely strong mutagens which have their effect to the DNA within the cell. Alkyl groups can be added to the nucleotides at various positions, in particular the O6 position of guanine. They tend to cause multiple mutations near the replication fork. The modified steric relationships caused by the adding of the alkyl group may result in the fragility of the DNA chain. Their impact can occur also on the next

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



cell cycle, for example defective repair happens, causing altered-base pairing (transition).

A variety of irradiation is also suitable to generate mutations. The higher-energy radiations such as X-ray and gamma radiation require expensive equipment and safety devices, and are not really suitable for routine use in microbiological laboratories, despite the fact that they have been used in the past. In addition, they produce excessive amounts of gross chromosomal damage which cannot be easily repaired by the microorganisms. In contrast, the ultraviolet (UV) irradiation can be easily controlled (although the eyes and skin protection is required), and requires relatively low cost equipment.

The main effect of UV radiation in the production of pyrimidine dimers (commonly referred to as thymine dimers, although the effect may also occur if the cytosine). UV causes covalent bond between two neighbouring pyrimidine bases in a strand of the DNA. These pyrimidine dimers do not serve as a template during replication; therefore they may be lethal for the cell, unless the cell is not able to repair the damage. An attempt to rectify that UV irradiation caused damage to the cell, can lead to mutagenic effects, as against the pyrimidine dimer the DNA polymerase enzyme randomly incorporates nucleotides, ultimately creating base pair substitutions.

We may talk about mutations only when a change occurred in the genetic material is further inherited. If the damaged DNA can be corrected before cell division, then no mutation is produced. Several error correction systems are present in the cells. Most of them work without error, others are error-prone. The error-correction systems can be categorized as follows:

- Direct reversal: In this case the enzymatic improvement of the chemically modified, but still recognizable bases occurs. In fact this is the reverse of the chemical reaction causing the mutation.
  - Photoreactivation: The best protection against the damage caused by UV irradiation is photoreactivation. This is catalysed by an enzyme (photolyase), which in the presence of visible light (photon needed) cleaves the covalent bonds linking pyrimidine residues, thus restoring the original situation. Therefore, when it is desired to induce mutagenesis in a laboratory using UV light, it is necessary to block the light from the cultures (e.g., by wrapping the bottle).
  - Alkyltransferase: At the formation of guanine receiving an alkyl group in the O-6 position (for example, when using methylnitrosoguanidine or EMS) the alkyltransferases are able to directly remove the group. The *E.coli* methyltransferase adds methyl group to the -SH group of the cysteine (S-methylcysteine arises), thus guanine are located again instead of the damage. In the reaction the enzyme (methyl acceptor protein) is permanently damaged.
- Excision repair: The so-called excision repair system cuts out the damaged section and based on the undamaged strand restores the original order of the nucleotides.
  - Base Excision Repair: A single erroneously incorporated base is cut off, which is refilled by the DNA polymerase using the complement strand as template.
  - Nucleotide Excision Repair: In this case the fault causing short stretch of single-stranded DNA (e.g. a thymine dimer plus the preceding and following few nucleotides) is cut off. The oligonucleotide sequence which has been cut off is replaced by a DNA polymerase and the fragments are connected by DNA ligase.
  - Mismatch repair: It repairs defective nucleotide pairs arising during the DNA replication and the subsequent recombination
- Repair of double-strand: The most dangerous is when the double stranded DNA is damaged. The repair of injuries requires





recombination mechanisms and may be required error-prone repair, too.

The SOS repair (error-prone) is known only in prokaryotes (although there is expected to exist similar mechanisms in eukaryotes as well), and it operates only in extreme cases when the survival of the cell is at stake. In the case of a heavily damaged DNA the goal is not necessarily the fault recovery, but the creation of a relatively functional DNA, providing a chance to survival. The replication complex is supplemented by subunits, which “forces” the “transcription” over the damaged area, despite the fact that the template sequence is not suitable for correct copying. Therefore, a large number of errors accumulate. The *Escherichia coli* SOS system is composed of at least 20 genes (*lexA*, *recA*, *uvr* genes).

During the induction two gene products are produced, namely: the RecA and LexA. The LexA-protein functions as the repressor of the SOS response genes (including *recA* and *lexA*). As a result of the damage of the DNA the RecA is linked to the single-stranded DNA. To connecting to the DNA induces conformational changes in the RecA, resulting in an ability to bind to the LexA. The LexA protein is able to cleave itself, but only if the RecA protein binds to it, as the RecA protein has coprotease functions, thus stimulating the independent proteolysis of the LexA. The proteolysis of the LexA allows the expression of the SOS genes, particularly the *umuC* and *umuD*, which are involved in the mutagenesis. (Those strains which lack these genes have not only increased UV sensitivity but also are not susceptible to UV-induced mutagenesis). The UmuCD complex is able to operate as DNA polymerase, taking the role of an ordinary polymerase. The low specificity of polymerase UmuCD allows a resumption of DNA synthesis.





## 9 TOOLS OF GENETIC ENGINEERING TECHNIQUES I (RESTRICTION AND MODIFICATION ENZYMES, ESSENTIAL MOLECULAR CLONING)

Different enzymes are used in genetic engineering. One of them is the group of restriction endonucleases. They were discovered in prokaryotes in the 1960s wherefore Werner Arber, Hamilton Smith and Daniel Nathans Othanel received the Nobel Prize in Medicine or Physiology in 1978. Nowadays, restriction enzymes are very important tools in modern molecular genetic research and are widely available commercially.

The basic function of the restriction enzymes is the protection against pathogens. The restriction enzymes of bacteria are capable to cleave the DNA of bacteriophage thereby prevent their growth in the host cell. However, the restriction enzymes cannot distinguish between the cell's own DNA and the DNA of phage intruders, consequently the host cell has to protect its own DNA against its own restriction enzymes. This protection is achieved by modifying enzymes. The restriction enzymes have suitable modifying enzymes as pair which have the same recognition sequence. The modifying enzymes modify chemically the cell's own DNA recognition sequences. The resulting modified sequences cannot be detected and cannot be cut by restriction enzymes. The host cells use typically DNA methylases for the modification. The methylases indicate with methyl groups the cell's own DNA at the recognition sites of the restriction enzymes, thus that is distinguishable from the phage DNA.

Most restriction enzymes are homodimeric protein, which are composed of two identical polypeptide units. Both units recognize and cleave one of the two strands of the DNA which causes the break of the double-strand of the DNA. The restriction endonucleases can be classified into four groups:

### Type I:

- cutting of the DNA happens outside the recognition site
- ATP and S-adenosyl-L-methionine are necessary for the operation
- cleavage and methylation features are also available

### Type II:

- cutting of the DNA happens always in the same way at the specified location within the recognition site or very close to it
- magnesium ion is needed, but ATP is not required for the operation
- there is only endonuclease activity (type II enzyme lacks methylation features)

### Type III:

- cutting of the DNA happens near to the recognition
- only ATP is required for the operation
- type III restriction endonuclease forms complex with methylase

### Type IV:

- it cuts modified ( i.e., methylated, glucosyl hydroxymethylated or hydroxymethylated) DNA





The molecular biological application of I and III types restriction enzymes is not typical. However, type II is essential for many kinds of jobs. Many type II enzymes cut the DNA in a staggered manner, which leads to the formation of short, single-stranded segments called sticky ends. There are a wide range of uses of sticky ends, for example those are applied during cloning. Other restriction enzymes cut straight across the double DNA strands producing blunt ends.

Most restriction endonucleases recognize a 4-6 (maybe 8-10) nucleotide length sequence. The recognition sequence is palindromic. This means that the base sequence of the two strands of the DNA shows central symmetry. If the restriction enzymes have identical recognition sequences and cleave the DNA at the same location, then they are called isoschizomers. Neoschizomers recognize the same nucleotide sequence as their prototype but cleave at a different site.

1 U is the amount of enzyme which is able to completely digest 1  $\mu\text{g}$  DNA in 60 minutes under optimal conditions. Temperature optimum of the restriction enzymes is generally 37 °C. The restriction enzymes are heat-sensitive, thus they are stored in glycerol-containing solution at -20 °C. The amount of the needed enzyme has to be in the range 1 to 5 U/  $\mu\text{g}$  DNA, but it is possible to apply them in a slightly larger quantity. The DNA sample should be cleaned before the digestion to reduce the presence of contaminants. For the functioning of the enzymes  $\text{Mg}^{2+}$  is required, and the enzymes may have different ionic strength and pH optimum.

The identification of the restriction enzyme sites of a DNA fragment gives the restriction map of this DNA fragment. In this case, the unknown DNA sequence is digested with restriction enzymes (singly or in combination) and the resulting fragments are separated by agarose gel electrophoresis.

Electrophoresis is a process that separates fragments in electric field. The pace of progress depends on the charge, the size and the shape of the fragments. During the gel electrophoresis the separation of molecules takes place in a porous gel. The gel is made of refined agar-agar (a polysaccharide mixture that is derived from algae) called agarose. The larger the molecule is the more it is prevented its progress in the cross-linked agarose. Consequently, the concentration of agarose must be selected accordingly to the size of the molecules to be investigated. In electric field the nucleic acids move towards the anode due to their negatively charged phosphate groups. Small molecules obviously move faster, while longer fragments are slower. The DNA has to be dyed in order the fragments to become visible. The ethidium bromide is a mutagenic dye which bounds to the double-stranded DNA and fluoresces the effect of UV light excitation. It can be added directly to the sample before its application to the gel or to the gel during gel preparation (in this case, the gel must be cooled and the dye has to be mixed with a liquid gel below 60 °C due to its heat sensitivity). After the electrophoresis the bands can also be dyed, if the gel is soaked in a solution containing ethidium bromide. For further use cutting out and purification of certain bands from the gel is possible as well.

The molecular cloning is propagation of a recombinant DNA containing a specific insert into a suitable host (recipient cell). Natural or synthetic DNA pieces are created by genetic engineering methods and they are incorporated into simple and small genetic elements i.e. DNA vectors. The vector may be plasmid, cosmid, phage or artificial chromosome, which may be introduced into the host cell. The vector is able to be multiplied in the host. If the insert DNA is carried in the vector, it is replicated together with the vector.

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



Cloning usually follows the next steps:

- (1) The first step is to isolate the corresponding DNA sequence. We may use total genomic DNA, reverse-transcribed DNA from RNA template, gene or genes amplified by PCR (polymerase chain reaction) technique, or *in vitro* artificially produced synthetic DNA.
- (2) The recombinant DNA construct is created by the ligation of the DNA sequence to be cloned (insert) and the vector (carrier – necessary for amplification of the insert in the host cell) DNA. This requires cleaving the insert DNA and vector DNA with restriction endonucleases, and then the resulting DNA pieces can be linked by DNA ligase. The DNA ligases establish phosphodiester bond between the 5' phosphate and 3' hydroxyl groups of the double stranded DNA. The advantage of blunt ends is that the pieces produced by different restriction enzymes can be matched. However, unwanted products may be formed, therefore the ligation efficiency is low. In contrast, the sticky ends can only be connected if they are complementary to each other.
- (3) For the introducing of the recombinant DNA into the recipient cell several methods have been developed: transformation of bacterial cells, transfection of eukaryotic cells, infection, electroporation and gene gun. In practice, a mixture of the recombinant constructs are generated. Some cells contain the cloned gene and others do not.

As only a small part of the recipient cells takes up the vector in a stable way, we have to identify the colonies of the transformed cells. The transformed cells to be identifiable the cloning vector must carry a suitable selection property such as a gene (marker). The most commonly used selectable markers confer resistance to antibiotics (e.g. ampicillin, chloramphenicol, and tetracycline). In this case, the cells were plated on agar containing antibiotics. Only those cells can grow which have been taken up the vector and carry the resistance gene.

The selection of the cells containing recombinant vector can be done by several methods:

- Insertion inactivity: It is used when there are two different resistance genes in the vector as well. The insert is inserted into one of the resistance genes and the successful insertion results the loosing of the resistance.
- $\alpha$ -complementation: The vector carries a short segment of *E. coli* DNA, a portion of the  $\beta$ -galactosidase gene (*lacZ*), which encodes the first 146 amino acids of the N-terminus of the enzyme. The recipient cell encodes the C-terminus of the  $\beta$ -galactosidase. The N-terminal and C-terminal of the  $\beta$ -galactosidase are not active separately but those are able to associate to form the active enzyme, this is called  $\alpha$ -complementation. In the medium containing a chromogenic substrate (for example: X-gal) cells encoding the active enzyme become colourful. The enzyme is induced by IPTG.
- Restriction digestion: The analysis of the restriction pattern of the isolated plasmid DNA can be effective.
- PCR: Using the plasmid DNA as template for polymerase chain reaction the presence of the insert can be justified with appropriate primers.
- Plasmid DNA sequencing.







## 10 TOOLS AND TECHNIQUES OF GENETIC ENGINEERING II (PLASMIDS, SHUTTLE AND EXPRESSION VECTORS, PLASMIDS EXPORT, OTHER VECTORS: YAC, BAC)

The requirements of the recipient host cells are the following: they must be easy to grow; growing fast in low-cost medium; to be genetically stable culture, which has a set of enzymes required for replication of the vector; non pathogens.

Host cells may be prokaryotic or eucaryotic. The most widely used prokaryotic host is the *Escherichia coli*. The great benefit is that a lot of molecular cloning methods are carried out in *E. coli*. Many, many tools are available for the cloning and genetic modification of this bacterium. On the other hand it is a pathogen therefore the work needs caution. Another disadvantage of the *E. coli* and the Gram-negative bacteria is the second membrane which greatly hampers the secretion.

Of course, this problem can be solved by using Gram-positive prokaryotic host. Many plasmids and phages are available for *Bacillus subtilis*, although the transformation technology is not so well established then in *E. coli*. In addition, the major disadvantage of *Bacillus subtilis* is the plasmid instability. It is often difficult to sustain the replication of the vector and the insert DNA; the bacterium may suddenly lose it.

The main incompleteness of prokaryotic host cells is the lack of the systems responsible for the modification of eukaryotic proteins. This problem can be solved using eukaryotic host cells. Eukaryotic cloning technology is well developed in *Saccharomyces cerevisiae*. Eukaryotic cells possess the necessary RNAs and post-translational modifying complex systems for eukaryotic protein synthesis. Thus, these are not required to incorporate the vector or host cell, as it must be done when prokaryotic hosts are cloned with eukaryotic insert. Cloning of mammalian cells is also possible. The cultured mammalian cell lines are major contributors to the investigation of various infectious and genetic diseases, and cancer research. They are maintained as the microbe cells, however, they are more sensitive. The insect cell lines are easier to handle.

Different types of vectors are used for cloning:

- Plasmid: Plasmids are the simplest bacterial origin vectors; Plasmids are extra chromosomal, double-stranded, closed ring (circular) DNA molecules which are present in many bacterial species. The size of their DNA is between 1,000 to 20,000 base pairs (bp). Plasmids often contain genes which encode enzymes that provide a selective advantage to the host cell. Such an advantage can be resistance to certain antibiotics or in other cases the synthesis of specific antibiotics or of different toxins. Some restriction-modification enzymes are also encoded in plasmid systems. The plasmids replicate independently of the host cell genome. An average plasmid vector can accommodate a 4000 bp length insert.
- Shuttle vectors (generally, plasmids): Shuttle vectors are capable to replicate and stably survive in two or more unrelated host organisms. Genes carried by the shuttle vectors are able to move between the different organisms (*E. coli* - *B. subtilis*, *E. coli* - yeast, *E. coli* - mammalian cell). Their importance lies in the fact that the







DNA cloned to an organism can replicate in another organism without any modification of the vector. This vector should contain two origins of replication, such as a bacteria or yeast.

- Expression vectors (generally, plasmids): The expression vectors contain regulatory sequences which allow the regulation of gene expression following the cloning. Such a regulatory sequence can be a promoter. In general, the aim is to maximize the yield. If the promoter is strong, it will be a high level of transcription. If the product is toxic a small quantity is sufficient and the control is very important; in this case a weak promoter is useful, such that the gene is written in a minor extent, or not at all.
- Phage vectors: Bacteriophages are viruses that infect bacteria. The advantage of the phage vectors is that by their infective properties they are capable with 100 % efficiency to introduce the foreign DNA into the host cell. A typical bacteriophage consists of an external capsid protein (the coat of the virion) and the genetic material inside. Their genome may be DNA or RNA that may be single-stranded or double-stranded. In gene technology the icosahedral  $\lambda$ -phage and the filamentous M13 phage are mostly used to create vectors. The double-stranded, linear DNA of the  $\lambda$ -phage encodes proteins necessary for the coat structure, the packing and the reproduction. The  $\lambda$ -phage is capable of two types of life cycle a lytic and a lysogenic one. During the lytic life cycle the phage is replicated multiple times in the *E. coli* cell, after that following the cell lysis the newly formed phages are able to infect further cells. During the lysogenic life cycle its genetic material is integrated into the host cell chromosome and replicates with it. The latent phage further inherited by the cell division is called prophage. The middle third region of the DNA is not necessary for the lytic process, so that can be substituted by the foreign DNA. On average, the  $\lambda$ -phage can accommodate 20000 bp insert, which can be packaged in vitro using the phage *cos* region and proteins. The non-lytic M13 phage can also infect *E. coli*. The single-stranded phage DNA forms double-stranded DNA during the replication, which becomes again single stranded during the transfer via pili and thus it infects further. In laboratory practice, the double-stranded replicative form is used as a cloning vector.
- Cosmids: The cosmids are hybrids of plasmid vectors and phage vectors. Specifically a cosmid is plasmid vector in which the *cos* region of phage is integrated. Thus, the cosmid is maintained as a plasmid in the cell while the packing of it into phage proteins enables the efficient phage infection of the host. Those are capable of receiving 40-60000 bp insert.
- Artificial vectors: With the development of genetic engineering very large DNA insert carrying vectors are necessitated for genome-wide analysis. For this purpose artificial chromosomes were developed which are suitable for construction of DNA libraries. The first artificial vector was the PAC (*P1-derived Artificial Chromosome*), which was produced from the P1 bacteriophage of *E. coli* and can accommodate approximately 100 to 300 Kbp insert. BAC (*Bacterial Artificial Chromosome*) is the modified version of the episome of F-factor from *E. coli* which is capable of carrying 350 Kbp insert. The YAC (Yeast Artificial Chromosome) has been designed from the genome of *Saccharomyces cerevisiae* and it can store up to Mbp size fragments as well. The mammalian cell-derived HAC (Human Artificial Chromosome) is the most complex type of vector and a promising tool for gene therapy as a "gene supplier" molecule. HAC vector can be created from an existing chromosome or de novo. The carried insert has no size limit.





Gene transformation can be carried on by different methods.

Transformation is the most common method for the introduction of the vectors into the bacteria cells. In this case, the cells take up DNA molecule (e.g. plasmid) from their environment. This phenomenon occurs even under natural conditions. In practice this can be achieved by generation of competent cells. Firstly, the cells are incubated in solution containing divalent cation (often  $\text{CaCl}_2$ ), which promotes the DNA binding to the bacterial cell wall and then competent cells are generated using heat shock. The cells are incubated on ice with the DNA, and then briefly heat-shocked (e.g., at  $42^\circ\text{C}$  for 30–120 seconds). Meanwhile, the cells take up the DNA (the exact mechanism is not understood). This method works very well for circular plasmid DNA.

The transfection is basically the same as the transformation however the target cells are mammalian cells. In this case, the negatively charged DNA is bound to a carrier, and this complex is introduced into the cell. Carriers such as cationic lipids package the DNA into liposomes which fuse with the cell membrane and introduce the vector into the cytoplasm. The transfection is also possible with cationic polymers (for example: DEAE-dextran, polyethyleneimine) or using calcium phosphate as well.

The vector can be absorbed by a bacteriophage or other virus-mediated gene transfer during infection.

An elemental carrier made from metal (e.g. gold) is enveloped with the DNA, then it is shot into the cell by high pressure (helium gas operated) gene gun device.

Finally, in case of electroporation, the cells are placed into solution and using an electroporator device electromagnetic field is formed in the solution. Due to the changed field strength transition holes are formed in the cell wall. The inclusion of the vectors occurs through the holes. The effectiveness of the method can be increased, if the vectors are present in high concentration in the solution. Although, many cells die during the procedure, electroporation is one of the most effective gene transfer method.





## 11 TOOLS AND TECHNIQUES OF GENETIC ENGINEERING III (NUCLEIC ACID HYBRIDIZATION AND SOUTHERN BLOT)

If the DNA is denatured i.e. the double strands are separated, the single DNA strands are able to form hybrid double strand with other single -stranded DNA or RNA molecules in accordance with the rules of complementarity. This is known as nucleic acid hybridization. If the DNA or RNA hybridize to the target, DNA sequence is appropriately marked (radioactively, enzymatically or fluorescently), then the method will be capable of detecting, identifying and characterizing DNA segments. The labelled nucleic acid fragment is called test.

The method for the analysis of the DNA was developed by Southern, who was rewarded with Nobel Prize in 1975. The method was named after him for Southern blot. Southern digested the DNA with restriction endonucleases, after that he separated the DNA fragments using agarose gel electrophoresis and transferred those from the gel to a cellulose nitrate filter, and then he hybridized the DNA fragments with radiolabelled ribosomal RNA fragments from *Escherichia coli*, *Xenopus laevis* and a wide range of mammals.

The method for the detection of RNA sequences was introduced later and was called Northern blot.

In the event that you are looking for fast detection of the target sequence and do not want to gain information about the target nucleic acid size, it is possible to skip the separation by size and the unseparated nucleic acid sample can be applied directly to the membrane. Dot Blot and Slot blot differ only in the form of patches. Small, circular spots are made in the case of dot blot, while slot blot occurs when rounded rectangular patches are prepared.

The following steps are followed during Southern blot:

1. Digestion: For the identification of the query gene or part of the gene the genomic DNA is cleaved with restriction endonucleases.
2. Separation by size: Separation of gene fragments or PCR products is done by agarose gel electrophoresis.
3. Membrane: The mechanical properties of the gel does not allow the direct analysis of the fragments by hybridization, therefore the resulting DNA pattern must be transferred to an easy to use membrane surface. This is the blotting step. Previously, a nitrocellulose membrane is used for this purpose. It binds well the nucleic acid, however because of its limited mechanical stability multiple washing and rehybridization of the membranes were not possible. These disadvantages are rectified by introducing a nylon membrane, although its nucleic acid binding ability is less than the one of the nitrocellulose membrane. Consequently the best membrane material is the chemically modified nylon, which carries a positive charge so its mechanical stability and nucleic acid binding ability are proper.
4. Depurination, denaturation: The bands migrate at different speeds due to their size difference. Because of the long transfer time leaching of the nucleic acid can occur, despite of the membrane binding ability. While the bigger bands will be clearly visible, the smaller ones





fade away. To solve the problem depurination was introduced. If the DNA is incubated in an acidic medium, its purine bases break off randomly and the disruption of the sugar -phosphate backbone occurs. Therefore, the soak of the gel in an acidic solution leads to the fragmentation of the DNA into smaller pieces. The neutralization of the gel in an alkaline medium is important after the depurination step. The resulted DNA fragments which are present in the respective bands migrate to the membrane in the original place according to their original size, however with different speed according to their reduced size. Denaturation of the double-stranded DNA is necessary for the subsequent hybridization when the labelled nucleic acid probe binds to the single-stranded DNA.

5. **Transfer Methods:** The classical method uses capillary force to move the liquid. The pretreated gel is laid on an appropriate solution-soaked filter paper and then membrane is placed on the surface of the gel. The whole is covered with a thick stack of filter paper which soaks up the moisture. While the paper soaks the liquid through the gel and the membrane, the DNA also moves together with the solution, however when it reaches the membrane, it binds. The method is simple, cheap and reliable. While the filter paper layer is saturated with a buffer solution, it continuously loses absorbency and the transfer will stop automatically. The only drawback is the large demand for time (approx. 10 hours overnight). The transfer can be accelerated when the flow of the liquid is pressure or suction-assisted. The migration of the DNA can be made faster by applying electrical voltage.
6. **Fixation:** The pattern of the nucleic acid bands has to be fixed by covalent bonds. Previously, this was solved by keeping the membrane for 2-3 hours at 80 ° C after its drying. Due to the flammable nature of nitrocellulose membrane this could be carried out only in vacuum exicator. The vacuum exicator is not necessary if nylon or modified nylon membrane is used. Another way of fixation is the UV light treatment and the development of photochemical bonds. The surest solution is the combined use of the two methods. After fixation, the membrane can be stored for a long time (several days) or used immediately. Before further use, the membrane must be wetted. The best way of this is to place the membrane on the surface of 6xSSC solution and wait until it absorbs moisture.
7. **Prehybridization:** In order to prevent nonspecific binding of the probe to the reactive binding-sites of the membrane, prehybridization of the membrane is required. The membrane is incubated in high density prehybridization solution containing indifferent DNA. The indifferent DNA can be extracted genomic DNA from the salmon or herring sperm or cheaper DNA derived from chicken blood. The prehybridization is carried out in the same instrument and under the same conditions as the hybridization. Shaking Water Bath can be apply. The membrane and prehybridization solution is placed nylon bag (the bag is) and thus is inserted to the bath. Safer and more economical is the insertion of the membrane and the pre-hybridization solution into a lockable glass tube and using a hybridization oven which is capable of receiving and shooting for the tube. In the hybridization oven the tubes revolve continuously, so despite the lower volume, the continuous washing of the membrane is achieved. The prehybridisation takes at least 3 hours.
8. **Hybridization:** DNA, RNA and oligonucleotide probe which have significant degree of similarity to the nucleic acid sequence may be used for the hybridization. The probe can be labelled radioactively, fluorescently or enzymatically. After denaturation the labelled probe is added to the pre-hybridization solution, and then the solution is poured on the surface of the membrane. The hybridization usually takes a few hours. The circumstances of the





hybridization depend on the relationship between the probe and the target sequence. It is a general rule that the temperature of the hybridization is lower than the melting temperature ( $T_m$ ) of the duplex formed by the probe and the target. If the size of the sequence identity is not exactly known, performing of the hybridization at lower temperatures is advisable, and then it is advisable to carry out hybridization at lower stages of severity, and then the weakly bound probes can be washed from the membrane by increasing stringency washes. The melting point of double-stranded nucleic acid (DNA or RNA homoduplex or DNA-RNA heteroduplex) and consequently the stringency of the hybridization can be controlled by changing the salt concentration of the solution and by adding organic solvent to the solution. High salt concentration and absence of organic solvent lead to mild hybridization conditions; in contrast stringent hybridization conditions can be provided in the presence of low salt concentration without organic solvent (e.g. dimethylformamide). There is linear relationship between the degree of rigor and the temperature of the solution. The most stringent hybridization conditions only allow for the connection of fully complementary sequences. In this case, the use of temperatures close to the melting point is expedient.

9. **Washing:** The surplus probe which is not bound to the target must be removed. The first step is the separation of the hybridization solution (it can be reused) from the membrane. Then, the membrane may be flushed with concentrated salt solution, and after that the washing process may be started. It is advisable to start the washing procedure under mild conditions (high salt concentration, low temperature) and then gradually increase the stringency of the wash.
10. **Detection:** Previously only radioactive labelling was used, mostly  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$   $\beta$  emitting isotopes and the detection of these was carried out by autoradiography. Nowadays, using non-radioactive labelling methods is very often, such as fluorescence, indirect and enzymatic labels. Due to its low sensitivity, fluorescent label is used only in rare cases. Particularly favored is the method called ECL which is based on chemiluminescence and is catalysed by an enzyme. The generated lights during the process can be fixed on disk images.

Northern blotting is suitable for direct examination of total RNAs or mRNAs, detection of specific mRNAs and determining their size. There are only a few differences between the construction of Northern blot and Southern blot:

- It is important that throughout the period of study RNA-ase free conditions have to be provided, since the RNA is very unstable.
- There is no need for restriction digestion, whereas the size of the corresponding mRNA strands to be tested is appropriate.
- Separation of the mRNA molecules by size happens on denaturing conditions. Denaturant is given to the agarose gel, therefore there is no required additional denaturation step. In practice two denaturants are used: glyoxal dimethyl sulfoxide (DMSO) or formaldehyde. Execution of DMSO treatment is technically more difficult because, during the electrophoresis the buffer (10 mM phosphate buffer pH 7) has to be continuously circulated in order that the pH remains stable, since glyoxal dissociates from the mRNA already in a slightly alkaline medium (pH > 8.0). However, it gives a sharper picture than the toxic formaldehyde. The dyeing of the RNA is done after the electrophoresis, because glyoxal can react with the ethidium bromide. In case of formaldehyde treatment the DNA runs





## DEBRECENI EGYETEM

faster than the RNA, therefore, DNA standard cannot be used for size determination.

- The prehybridization, hybridization and detection is made as described above for Southern blots.







## 12 SEQUENCING, SYNTHESIS AND AMPLIFICATION OF DNA

Several methods are developed for the determination of DNA base sequence. The first method has been developed by F. Sanger in 1975; the other method was published in 1977 as a result of AM Maxam's and W. Gilbert's work. The Maxam-Gilbert technique initially enjoyed great enthusiasm, but it is a very complex method which additionally requires the use of radioactive isotopes and other hazardous chemicals, thus it was soon forgotten and replaced by Sanger's method.

### Maxam-Gilbert method:

Firstly the DNA is cut by restriction enzymes. The resulting small fragments shall be marked with  $^{32}\text{P}$ -containing phosphate group. Four parallel reactions are necessary. In each reaction the DNA is digested at a specified nucleotide or nucleotides. The resulting even smaller fragments are separated by electrophoresis, and then it is possible to identify the fragments on the basis of their chain length. The base and nucleotide sequence deduced from the position of the fragments in the gel in all four reactions.

### Sanger method:

- **Labelling:** The first step of the method is making single stranded DNA, after that suitable, labelled oligonucleotide primer shall be hybridized to the 3' end of the single-stranded DNA. The labelling can be radioactive or fluorescent. Subsequently, the DNA polymerase is able to build up the complementary DNA strand. For the chain extension the mixture divided into four parts and the DNA building blocks the deoxynucleoside triphosphates (dNTPs: dATP, dGTP, dTTP, dCTP) are added into them. Each different mixture contains different modified dNTPs (ddNTP).
- **Chain termination:** If a dNTP is modified to include the 3' position which does not contain a hydroxyl group (ddNTP: 2'- 3'-dideoxy- nucleoside triphosphate), after its incorporation into the DNA the synthesis of the strand cannot continue.
- **Detection:** The appropriate rate of dNTPs and ddNTPs ensures that all possible ends statistically occur. The method executed with four different ddNTPs results four independent reactions. When the products are separated by polyacrylamide gel electrophoresis four separate gel lanes are obtained. After autoradiography or fluorescent detection the sequence becomes readable.

For example, during the PCR reaction, adenine is incorporated into the new strand opposite thymine. If the reaction mixture contains both dATP and ddATP, the incorporation of dATP and ddATP opposite the thymine bases of the template DNA are equally possible. If dATP is incorporated, the synthesis of the new strand may continue. Otherwise, the incorporation of ddATP stops the synthesis. During the reaction, fragments with various lengths are formed. The lengths of these reflect the distance (position) of the thymine bases from the 5'-end of the primer.

The advantages of the method are its simplicity, accuracy and automation. The automatic Sanger method is considered as the first generation sequencing method. During the process four different colours of fluorescent dye (fluorophore) are connected to the four kinds of modified nucleotides (ddNTPs) instead of the primers. The synthesis of the complementary

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**





DNA strand is carried out in one reaction. The reaction leads to fragments with different lengths, and all of them have fluorescently labelled end. The products are loaded into one gel well and separated by polyacrylamide gel electrophoresis. The bands are detected by fluorescence spectroscopy, which is able to detect each of the four distinct fluorophore. The template DNA sequence is determinable from the bottom toward the top of the gel.

In the fully automatic method, the products are separated by automated electrophoresis (capillary electrophoresis) and the bands are detected by fluorescence spectroscopy. In the capillary electrophoresis there is capillary (50 to 70 cm long, 50 to 100  $\mu\text{m}$  inner diameter) which is filled with cross-linked gel matrix. The samples are loaded into the capillary, and the fragments migrate due to electrical voltage (DNA migrates towards the anode). Small fragments move faster because of the filter effects of the gel matrix. At the end of the capillary the detection of the four type of fluorescence signal is continuous. The result of the continuous detection is a chromatogram attached by computer. The incorporated nucleotide analogues can be identified on the basis of the colour of the fluorescence by the application of corresponding excitation.

### DNA synthesis

Already in the fifties Michelson and Todd were able to polymerize nucleotides. Firstly Khorana synthesized a 77 nucleotides length gene in 1970 and for this he was awarded by the Nobel Prize. The method developed by Khorana and his co-workers was a slow and laborious phosphodiester method. After that Letsinger developed the phosphor-triester method which is the currently used version of the phosphoramidite method. From the mid-80s automated oligonucleotide synthesizer were invented.

In the automated version of the phosphoramidite method, the in vitro DNA synthesis is a solid phase process in which the first nucleotide of the chain is fixed at the solid support (four kinds of support are required for the different nucleotides), which may be a porous glass or polystyrene-based resin. Attachment of each nucleotide is carried out by a synthesis consisting of four main steps. Oligonucleotides were synthesized in the 3' -5' direction. During the synthesis, the washings among the steps are very important. When the oligonucleotide has reached a sufficient length, it is cleaved from the solid support and purified.

Main steps are the following ones:

- Activation (de-blocking): To add the next base, the 4,4'-dimethoxy trityl (DMTr) group protecting the 5'-hydroxyl group must be removed. The ether bond of the acid sensitive DMTr can be decomposed due to mild acid hydrolysis. This causes the separation of the DMTr and the support-bound, protected deoxynucleoside.
- Coupling: The next base monomer cannot be added until it has been activated. This is achieved by adding tetrazole to the 5'-hydroxyl group of the free base. deoxynukleozid. For these step phosphoramidite is used because it is stable and it can be stored easily. The phosphoramidite is converted to tetrazole immediately prior to the reaction (A foszformaidit phosphite and a secondary amine is formed.)
- Block (capping): Acetylation of the free 5'- hydroxyls is important to avoid the formation of undesired sequences during subsequent

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME





cycles. (Since the connection is not perfect failure sequences are not created, but instead shorter oligomers are synthesized.)

- Oxidation: The phosphite is oxidizing to a phosphate group typical of natural nucleotides.

### **Polymerase Chain Reaction (PCR)**

The PCR method was developed in 1985 by Kary B. Mullis. For this discovery he was awarded a Nobel Prize in Chemistry in 1993. PCR is capable of the exponential amplification of a specific DNA fragment by in vitro DNA synthesis. The purpose of the procedure is to generate a particular DNA segment (e.g. a gene) in large quantities. During the PCR reaction DNA polymerase enzyme carries the DNA chain extension. Initially, the proteolytic fragment of *E. coli* DNA polymerase I enzyme was used in the PCR reaction. This is called Klenow fragment, which lacks exonuclease activity. The enzyme is heat sensitive therefore fresh enzyme had to be added to the mixture after every denaturation. Whereas the polymerization was carried out at low temperature (37 ° C) defective hybridization was frequent. However, in 1988 a new polymerase was introduced, which is meant to solve the above problems. The Taq polymerase derived from *Thermus aquaticus* is a thermostable enzyme: its half-life is 130 minutes at 92.5 ° C. The optimum temperature of its operation is 70-80 ° C.

Two artificially synthesized oligonucleotide primers are also required for the reaction, which hybridize to different strands of the template DNA (3' ends face each other). During the process the DNA region (target) which is terminated by the primers is multiplied. The template DNA may be of any origin. The building blocks of DNA (dNTPs) are required for the synthesis of the new DNA strand. The reaction mixture has to contain all four types of dNTPs in the same amount. MgCl<sub>2</sub> is also an important component of the reaction mixture because it is necessary for the functioning of the polymerase. The standard PCR reaction components and the quantities are summarized in Table 1.

**Table 1:** Components of the Polymerase Chain Reaction

Component	Concentration
Template DNA	10-100 ng
PCR buffer	1/10 of final volume
MgCl <sub>2</sub>	0,5-5 mM (typical 1,5 mM)
dNTPs	20-200 μM from each
Forward primer	0,1-0,5 μM
Reverse primer	0,1-0,5 μM
Taq polymerase	0,5-2,5 units
Sterile distilled water	To final volume



The synthesis of the new DNA strands is carried out in a programmable thermostat (PCR machine). The three reaction steps are repeated cyclically:

- (1) Denaturation: The double-stranded DNA becomes single-stranded because of the high temperature (94-97 ° C; 0.5-1 minutes).
- (2) Annealing: The oligonucleotide primers are hybridizes to the template DNA according to base pairing rules (that binds to the sense strand is called the forward or sense primer; the reverse or antisense primer adhesives bond with the antisense strand). The optimal annealing temperature is about five degrees lower than the melting temperature of the primers. The too low annealing temperature leads to non-specific hybridization. If the annealing temperature is too high, little amount of specific PCR product is formed or the reaction does not result any PCR product. This step takes 0.5-1 minutes.
- (3) Polymerization: The DNA polymerase starts the synthesis of complementary strand from the 3' end of primer. The optimum temperature for Taq polymerase is 72 ° C. The duration of the polymerization depends on the length of the synthesized product. The capacity of the enzyme is about 1000 bases/minute.

After 30-40 cycles of about  $10^6$  fold amplification is achieved, after that the exponential growth stops. The cycles are followed by a single final polymerization step (72 ° C, 5-15 min) in order to the synthesis of not fully polymerized sections be completed. After final polymerization the machine cools the reaction mixtures (4-8 ° C) to avoid the further process.





## 13 GENE MANIPULATIONS: SOME EXAMPLES OF APPLICATIONS IN BIOTECHNOLOGY AND BASIC RESEARCH

The genetic manipulation has widespread application in several fields of medicine and biology. Microbes are able to produce different valuable macromolecules (e.g. proteins, nucleic acids, carbohydrate polymers) or smaller metabolites (e.g. alcohol, citric acid, acetic acid). Moreover genetic manipulation provided the possibility of the production of commercially important products like therapeutic proteins, antibiotics, and small molecules that are isolated from plants or animals.

Recombinant human insulin was the first medicine approved by the FDA in the early 1980s, which was made via recombinant DNA technology and produced in *E. coli*. The bacterial expression systems are still preferable choice able to produce a recombinant human protein due to their simple and cheap cultivation. One-third of the approved 151 recombinant therapeutic proteins in 2012 were produced in *E. coli*. It is still the most attractive recombinant microbe, because well characterized genetics and it has been successfully used to express several different proteins.

Human insulin and human growth hormone are one of the clinically most important products of microbial biotechnology. Insulin is a peptide hormone, and used in diabetes mellitus treatment. Recombinant human insulin is produced either in bacteria, or in yeast. First the human insulin protein gene and plasmid vector must be isolated. Then the insulin gene is inserted into the plasmid, and the human DNA is now recombined with bacterial DNA. The bacterial cells must be transformed with the vector having insulin gene, and allow to grow and multiply. During this process, bacterial cells start to produce recombinant insulin. Finally human insulin produced by bacteria must be purified.

Other important field of microbial biotechnology are the development of systems for the identification of disease-causing organisms. Monoclonal antibodies are specific antibodies used in serological methods (e.g. detecting pathogens in food and water). They also can be used in cancer treatment due to their site specificity. Monoclonal antibodies are produced by identical immune cells. Recombinant antibody techniques based on to create antibody libraries, which are different in only few amino acids determining their specificity. *Aspergillus niger* is preferable to use for monoclonal antibodies production over the *E. coli*. Fungal cells produced antibodies extracellularly, which make protein separation process easier. Ward et al. (2004) fused the N-terminal of antibodies to glucoamylase, and a fungal endogenous protease (KexB) cleaved off the fusion during the secretion of the antibody.

The microorganisms are used for production of commercially important enzymes from several decades. Enzyme production could be improved by recombinant DNA technology with genetic modification to over-express the desired enzymes. Microbes used for ethanol fermentation are modified inserting enzymes from other microbes to utilize 5-carbon sugars, like xyloses releasing from the hydrolysis of woody biomass. An alternative approach would be integrating genes encoding the enzymes making up the ethanol fermentation pathway into cellulose degrading organism. Another strategy is to create microbes for producing novel or improved degradative enzymes, like cellulases or amylases. Those enzymes can be used to pre-treat cellulosic

TÁMOP-4.1.1.C-12/1/KONV-2012-0014

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

SZÉCHENYI 2020

Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE



feedstocks, or can be added to ethanol production process to enhance or accelerate biofuel production processes.

Enhancing remediation capacity is another aim for engineering microbes. Microorganisms may degrade or transform hazardous organic compounds into non-toxic substances during microbial remediation. The cleaning of toxic-waste sites from radiation exposed environments is a big challenge. The extraordinary DNA-repair processes of *Deinococcus radiodurans* enable it to survive the high-radiation exposed environments. The potential of cleaning up toxic-waste sites from radiation exposed environments has been increased with the help of genetic modification its genome with inserting specific enzymes from other organisms and make the *D. radiodurans* able to degrade toxic chemicals such as toluene. Genetically engineered microorganisms for bioremediation of oil spills have also been created. Different species of the Gram-negative *Pseudomonas* bacteria contain enzymes, which can break down various hydrocarbons. The oil-degrading enzymes are encoded by genes located on plasmids. Chakrabarty had been able to create a variant of *Pseudomonas* by combining plasmids with different degradative capacity into a strain of *Pseudomonas putida*. That “superbug” was capable of breaking down the constituents of crude oil 10–100 times faster than other non-genetically engineered independent strains. However the strain was never used, due to regulations and concerns.

Currently synthetic biology aims to create novel or synthetic microorganisms with enzymatic capabilities not found in the original host organism. This designing “optimal” organism would possess and use either combination of enzymes from different sources, or even completely new enzymes designed and created using protein engineering to have maximal catalytic activity.

Genetically modified microorganisms can be used a living sensor to different pollutants or toxic chemicals from different samples (soil, air, food etc.). Bacteria, for example, can emit a green fluorescent light when metabolizing TNT leaking from mines. The recombinant yeast-based estrogenicity assay, or recombinant yeast assay (RYA) is widely used for the detection of molecules capable binding to estrogen receptor (ER $\alpha$ ). The engineered yeast strain possesses that ER $\alpha$  receptor, and a reporter gene. The transcription of the reporter gene depends on the presence of an estrogen-like (ER $\alpha$  receptor binding) compound in the medium. Correct structure of the ligand-binding domain of the expressed human receptor protein is provided by the yeast folding and post translational processes, which is very similar to the vertebrate one. Moreover yeast has no endogenous system homologous to vertebrate nuclear receptors, which could interfere with the assay.





## 14 METABOLIC ENGINEERING OF MICROORGANISMS FOR THE PRODUCTION OF MOLECULES OF INTEREST FROM BIOMASS

Currently, microbes are used for commercial production of a wide variety of products such as antibiotics, enzymes, different metabolites, pesticides, fertilizers, feed additives, biopharmaceuticals, therapeutics, biopolymers and biofuels. Beyond substances produced by microbial cells, the microbial cells themselves are also important (e.g. yeast cells for baking, or brewing). The microbial products are originated from metabolic reactions of the microorganisms, but their overproduction have achieved with the modification of their metabolism.

It must be able to produce the substance of interest in high yield. Industrial microorganisms can be thought of as metabolic specialists, capable of synthesizing one or more products in high yield. Industrial microbiologists often use classical genetic methods to select for high-yielding mutant strains; their goal is to increase the yield of the product to the point of being economically profitable. The genetics of the producing organism needs to be well understood. After selection, the metabolic behaviour of the production strain may be far removed from that of the original wild-type strain.

A microorganism used in an industrial process must have special features. The organism must be capable of growth and product formation in large-scale culture. It means that an industrial microorganism should produce reproductive cell (e.g. spores) or some other to enhance inoculation into the large vessels; grow rapidly and produce the product in a relatively short period of time; able to grow in a liquid culture medium obtainable in bulk quantities at a low price. Fermentation medium can represent almost 30 % of the cost for a microbial fermentation. Therefore many industrial microbiological processes use waste carbon from other industries as major or supplemental ingredients for large-scale culture media. Currently, a wide range of industrial and agricultural by-products and waste materials are used as low-cost biomass resources (nutrients) for industrial fermentations. These include corn steep liquor (a product of the corn wet-milling industry that is rich in nitrogen and growth factors), whey (a waste liquid of the dairy industry containing lactose and minerals), and lignocellulosic biomass of different plants.

Further requirements for an industrial microorganism are not to be pathogenic, and should be amenable to genetic analysis. Because the yields necessary to make an industrial process profitable, typically demand is the selection of high yielding mutant derivatives of the original wild-type organism.

Antibiotics (substances produced by microorganisms that kill or inhibit the growth of other microorganisms) are produced by large scale industrial fermentations today. One of the major tasks is to isolate high-yielding strains from the natural (wild type) strains. Strain selection may involve mutagenizing the wild-type organism to obtain mutant derivatives with the ability of overproduce the antibiotic of interest.

Lysine is produced commercially by the bacterium *Corynebacterium glutamicum* for use as a food additive, being an essential amino acid for humans and domestic animals. The lysine production in bacteria is strictly regulated in cells. However, these regulatory mechanisms

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA



Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**





must be circumvented, as the overproduction necessary to make amino acids commercially from a microbial source. The production of lysine is controlled at the level of the enzyme aspartokinase, as excess lysine feedback inhibits the activity of this enzyme in *C. glutamicum*. Mutants lacking that feedback inhibition by lysine could be isolated. Typical mutants of *C. glutamicum* can produce over 60 g of lysine per litre in industrial fermentors, a concentration sufficiently high to make the process commercially viable.

Bio fuels include ethanol, biodiesel, and algal fuels. Biofuel are made from the fermentation of recently grown plant material rather than being of ancient origin (that is, fossil fuel). Ethanol is currently the most important global biofuel, but the feedstock used for ethanol production has been a major issue. For example, the increased demand for corn as a biofuel feedstock has driven up the price of human foods and livestock feeds in the United States. In other countries (e.g. Brazil), not only corn but also sugar cane, whey, sugar beets, wood chips and waste paper are used as feedstocks for the fermentation. For cellulosic materials, the cellulose must first be treated to release glucose, which is then fermented to alcohol.

Small metabolites are typically results of the operation of several enzymes, which are regulated in a coordinated manner. Pathway engineering is the process of assembling a new or improved biochemical pathway using genes from one or more organisms. Genetic engineering of bacteria is simpler than that of higher organisms; consequently most pathway engineering has been done with bacteria. Engineered microorganisms are used to make products (alcohols, solvents, food additives, dyes, antibiotics), to degrade agricultural toxic or undesirable materials (waste, pollutants, herbicides).

A suitable microorganism for converting the lignocellulose biomass into ethanol is still lacking. Lignocellulosic biomass contains complex carbohydrates, most significantly xylose. Some Gram-negative bacteria, like *Escherichia coli* and *Klebsiella oxytoca* are naturally able to use wide spectrum of sugars. These microbes have engineered toward selective production of ethanol and increased ethanol tolerance. *Zymomonas mobilis*, another Gram-negative bacterium produces ethanol at high yields, however only from glucose or fructose. Metabolic engineering of this organism is about introducing pathways for the fermentation of arabinose and xylose. Several metabolic engineering approaches have been employed in the yeast *Saccharomyces cerevisiae* (cultivated from a long time for ethanol production from sugars) to improve its xylose utilization.

Most efforts so far have modified and improved existing pathways rather than creating entirely new ones. The later one is the field of "synthetic biology", using genetic engineering to create novel biological systems out of available biological parts (from several different organisms). An ultimate goal of synthetic biology is to synthesize a viable cell from component parts. The A team placed a laboratory-synthesized chromosome of one bacterial species into the cell of a second bacterial species and got the synthetic chromosome to function normally and direct the activities of the recipient cell in 2010.





## 15 GENOME FUNCTION AND REGULATION (MICROARRAYS AND TRANSCRIPTOME, PROTEOMICS, METABOLOMICS)

The entire complement of genetic information (genes, regulatory sequences, noncoding DNA) of an organism is called its genome. Analyses of genomes (mapping, sequencing, analysing, and comparing genomes) make up the field of genomics. It includes the study of the transcription and translation of the genetic information at a genome-wide level (gene expression). Gene expression studies either focus on a single gene or group of related genes (traditional approach), or all or most of an organism's genes are examined in a single experiment.

The first genome sequenced was the RNA genome of the virus MS2 in 1976, and the first DNA genome sequenced was a small, single-stranded DNA virus,  $\phi$ X174, in 1977 by a group led by Fred Sanger. He has introduced the dideoxy technique for DNA sequencing. The genome of *Haemophilus influenzae* was the first cellular genome sequenced in 1995. At present sequencing of more than 2000 genomes of prokaryotes has either been completed or is in progress. New advances in DNA sequencing enhances the rapid growth of the number of sequenced genomes. Genome sequencing has now become routine, and the genomes of several different strains of the same bacterium have been sequenced in order to reveal the extent of genetic variability within a species.

Following the determination of the whole sequence of a genome, the next step is genome annotation, the conversion of raw sequence data into a list of the genes and other functional sequences in the genome. Most genes encode proteins, and in most microbial genomes, the great majority of the genome consists of coding sequences (microbial eukaryotes typically have fewer introns, than plant and animal genomes, and prokaryotes have almost none). Microbial genomes consist of hundreds to thousands of open reading frames separated by short regulatory regions and transcriptional terminators. Open reading frame (ORF) is a sequence of DNA or RNA that could be translated to give a polypeptide.

The smallest cellular genomes belong to parasitic or endosymbiotic prokaryotes, and range from 490 kbp of an Archaea to 4400 kbp for *Mycobacterium tuberculosis* (Bacteria). Their genomes are smaller than the largest known viral genome, that of Mimivirus with 1.2 Mbp. However the genome of some prokaryotes, like nitrogen fixing *Bradyrhizobium japonicum* (9.1 Mbp) are comparable in size to those of eukaryotic microorganisms. The haploid yeast genome contains 16 chromosomes ranging from 220 kbp to about 2352 kbp. The total yeast nuclear genome (excluding the mitochondria and some plasmid and virus-like genetic elements) is approximately 13,392 kbp.

An annotated genome sequence is simply a list of parts, and understanding how a cell functions, both gene expression (transcription) and the function of the final gene product must be investigated. The entire complement of RNA produced under a given set of conditions is known as the transcriptome (in analogy to the term "genome").

Measuring gene expression under different conditions may reveal a gene's function. Nucleic acid hybridization can be used in conjunction with genomic sequence data to measure gene expression by hybridizing mRNA to specific DNA fragments. This technique however has been

TÁMOP-4.1.1.C-12/1/KONV-2012-0014

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

SZÉCHENYI 2020

Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE



enhanced with the development of microarrays.

The microarrays used in genomics are small, solid supports (often called gene chips, Fig. 15.1) to which genes or segments of genes are fixed and arrayed spatially in a known pattern. The oligonucleotides (gene segments synthesized by PCR, or oligonucleotides designed for each gene) is attached to the solid support in a known pattern. Photolithography (used to produce computer chips), has been adapted to produce silica microarray chips 1 to 2 cm in size. Each DNA chip can hold thousands of different DNA fragments. In practice, each gene is often represented more than once in the array to provide increased reliability. The mRNA from cells grown under specific conditions are labelled and hybridized with the attached ssDNA on the chip. The mRNA sequences must first be amplified as isolated mRNA is in too low amounts for direct use for microarray analysis. Reverse transcriptase (RT) is used to generate complementary DNA (cDNA) from the mRNA. The cDNA may then be amplified either with DNA polymerase to generate multiple DNA, or with T7 RNA polymerase to generate multiple RNA copies (cRNA). The cDNA or cRNA is then applied to the array. The amplified mRNAs from the different cells or cell cultures are labelled differently during amplification (e.g. one with a green fluorescent dye, the other with a red one). Finally the chip is scanned: a laser excites each spot and the fluorescent emission gather through a photo-multiplier (PMT) coupled to a confocal microscope. Collected data are analysed by computer. Hybridization between a specific mRNA and a DNA segment on the chip indicates that the gene has been transcribed.

One chip may be used to assay expression of the *Saccharomyces cerevisiae* genome, holding the protein-encoding genes functioning 6200 ORFs) of *S. cerevisiae*. Consequently the global gene expression in this organism can be measured in a single experiment with applying the cRNA or cDNA derived from mRNA obtained from yeast cells grown under specific conditions and labelled with different fluorescent dyes e.g. green and red). Any particular cRNA/cDNA hybridizes only to the DNA on the chip that is complementary in sequence. A distinct pattern of hybridization is observed, depending upon which DNA sequences correspond to which mRNAs. The intensity of the fluorescence gives a quantitative measure of gene expression. This allows the computer to make a list of which genes were expressed and to what extent. Thus, using gene chips, the transcriptome of the organism of interest grown under specified conditions is revealed from the pattern and intensity of the fluorescent spots generated.

Gene chips may be used to compare expression of specific groups of genes under different growth conditions. The *S. cerevisiae* gene chip has been used to compare gene expression between yeast cells grow by fermentation and by respiration. Transcriptome analysis can reveal which genes are shut down and which are turned on when yeast cells are switched from fermentative (anaerobic) to respiratory (aerobic) metabolism or vice versa. Results show that yeast undergoes a major metabolic change during the switch from anaerobic to aerobic growth. Overall, over 700 genes are turned on and over 1000 turned off during this metabolic transition. For example a number of genes connected to ethanol production are strongly repressed, whereas citric acid cycle functions are strongly activated by the switch. Moreover, the expression pattern of genes of unknown function is also monitored during the fermentative to respiratory switch by using a microarray, yielding clues to their possible role. At present, no other available method can give as much information about gene expression as microarrays.

The comparison of genes in closely related organisms (e.g. pathogenic and non-pathogenic bacteria, normal or cancer cells) is another

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



important use of DNA microarrays. While DNA chips, containing the whole human genome, may be used in clinical diagnostics.

Several important data can be obtained by the transcriptome of an organism, however the number and types of proteins present in a cell change in response to different factors, like environment, and developmental cycles. The genome-wide study of the structure, function, and regulation of an organism's proteins is called proteomics. The term proteome in its narrower sense refers to those proteins present in a cell at any given time. However, in wider sense, refers to all the proteins encoded by an organism's genome.

Several methods are used in analysis of proteomes. The first major approach was the two-dimensional polyacrylamide gel electrophoresis (2D PAGE). This technique can separate, identify, and measure all the proteins present in a cell sample. In the first dimension, the proteins are separated by differences in their isoelectric points (the pH at which the net charge on each protein reaches zero). While in the second dimension, the proteins are separated by size. The presence of a particular protein under different growth conditions can be measured and related to environmental signals using 2D gels. Unknown protein can be eluted from the gel and sequence a portion of it to connect with a particular gene. Recently, eluted proteins have been identified by mass spectrometry, following a preliminary digestion. The information of a characteristic set of peptides may be sufficient to completely identify the protein. Partial protein sequence data may allow the design of oligonucleotide probes or primers to find the gene encoding the protein from genomic DNA by hybridization or PCR. Then, after sequencing of the DNA, the gene may be identified. High-performance liquid chromatography (HPLC) is increasingly used to separate protein mixtures nowadays. The stationary phase in the column separates dissolved sample proteins by variations in their chemical properties (size, ionic charge, or hydrophobicity). Collected fractions of the different proteins are digested by proteases and the peptides are identified by mass spectrometry.

The metabolome is the complete set of metabolic intermediates and other small molecules produced in an organism. The chemical diversity of small metabolites makes systematic screening technically challenging. Early attempts used nuclear magnetic resonance (NMR) analysis of extracts from cells labeled with  $^{13}\text{C}$ -glucose. However, this method has limited sensitivity, and has too low resolution for the complex cell extracts. The most promising approach to metabolomics is the use of mass spectrometry using extremely high mass resolution, allowing the unambiguous determination of the molecular formula of any small molecule. MALDI (matrix-assisted laser desorption ionization) is a version of mass spectrometry, where the sample is ionized and vaporized by a laser. The ions generated are accelerated along the column toward the detector by an electric field. The time of flight (TOF) for each ion depends on its mass/charge ratio - the smaller this ratio, the faster the ion moves. The detector measures the TOF for each ion and the computer calculates the mass and hence the molecular formula. The combination of these two techniques is known as MALDI-TOF. The same approach is used to identify the peptide fragments from digested proteins during proteome analyses. In this case, identifying several oligopeptides allows the identity of the parent protein to be deduced provided that its amino acid sequence is known.

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME

**SZÉCHENYI** 2020



MAGYARORSZÁG  
KORMÁNYA

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**



## Bibliography

Bickle TA, Krüger DH (1993). "Biology of DNA restriction". Microbiol. Rev. 57 (2): 434–50.

Desmond S. T. Nicholl: 2008. An Introduction to Genetic Engineering Third Edition. Cambridge University Press, New York, 336 p.

Friello D.A., Mylroie J.R., Chakrabarty A.M. (2001) Use of genetically engineered multi-plasmid microorganisms for rapid degradation of fuel hydrocarbons. International Biodeterioration & Biodegradation 48: 233–242

Garcia-Reyero N, Grau E, Castillo M, López de Alda MJ, Barceló D, et al. (2001) Monitoring of endocrine disruptors in surface waters by the yeast recombinant assay. Environ Toxicol Chem 20: 1152–1158.

Huang C.J., Lin H., Yang X. (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. Journal of Industrial Microbiology & Biotechnology 39: 383-399.

Jay D Keasling Zhou H, Cheng J-S, Wanga BL, Fink GR, Stephanopoulos G (2012) Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by *Saccharomyces cerevisiae*. Metabolic Engineering, 14:611-622.

Jeremy W. Dale and Simon F. Park: 2010. Molecular Genetics of Bacteria 5th edition. Wiley-Blackwell, Chichester, 386 p.

Kamionka M. (2011) Engineering of Therapeutic Proteins Production in *Escherichia coli*. Current Pharmaceutical Biotechnology 12: 268-274.

Klis FM, Boorsma A, De Groot PWJ (2006) Cell wall construction in *Saccharomyces cerevisiae*. Yeast 23: 185–202.

Li SC, Kane PM (2009) The yeast lysosome-like vacuole: endpoint and crossroads. BBA 1793(4):650-63.

Madigan, Michael T., Clark, David P., Martinko, John M.: 2011. Brock Biology of Microorganisms 13th Edition. Benjamin Cummings, San Francisco, 1043 p.

Molekuláris biológiai módszerek (Egyetemi jegyzet) 2004. Szerk. Dombrádi V. Debrecen: Debreceni Egyetem Orvos és Egészségtudományi Centrum 65-79 p.

Nadkarni, S.: 2009. Gene Manipulation. Swastik Publishers & Distributors, Delhi, 301 p.

Suda Y, Nakano A (2012) The yeast Golgi apparatus. Traffic. 13:505-10.

Ward M., Lin C., Victoria D. C., Fox B. P., Fox J. A., Wong D. L., Meerman H. J., Pucci J. P., Fong R. B., Heng M. H. et al. (2004) Characterization of humanized antibodies secreted by *Aspergillus niger*. Appl Environ Microbiol, 70: 2567-2576.





## DEBRECENI EGYETEM

Zhang, Rodriguez S (2011) Metabolic engineering of microbial pathways for advanced biofuels production. Current Opinion in Biotechnology 22:775-783.

<http://elte.prompt.hu/sites/default/files/tananyagok/biokemia/ch10s09.html>

[http://elte.prompt.hu/sites/default/files/tananyagok/biokemia\\_alapjai/ch19s05.html](http://elte.prompt.hu/sites/default/files/tananyagok/biokemia_alapjai/ch19s05.html)

<http://elte.prompt.hu/sites/default/files/tananyagok/gentechnologia/ch12s02.html>

<http://elte.prompt.hu/sites/default/files/tananyagok/prokariotak/ch03s05.html>

<http://ttk.pte.hu/biologia/genetika/atg/chap18/ch18a.htm#beg>

<http://ttk.pte.hu/biologia/genetika/atg/chap19/ch19c.htm>

<http://www.tankonyvtar.hu/hu/tartalom/tkt/oxford-typotex-biologiai/ch01s15.html>

<http://www.termeszetvilaga.hu/szamok/tv2007/tv0711/szabad.html>

**TÁMOP-4.1.1.C-12/1/KONV-2012-0014**

Élelmiszerbiztonság és gasztronómia vonatkozású  
egyetemi együttműködés, DE-SZTE-EKF-NYME



MAGYARORSZÁG  
KORMÁNYA

**SZÉCHENYI** 2020

Európai Unió  
Európai Szociális  
Alap



**BEFEKTETÉS A JÖVŐBE**