Organization of genomes, as well as genetic elements that regulate gene expression in pro- and eukaryotes

Lecturer: Senior Lecturer, Department of Molecular Biology and Genetics, PhD, Smekenov I.T.
Subject: Genetic engineering

(Lecture 6)

Lecture Goal:

To examine the organization of genomes and the genetic elements that regulate gene expression in both prokaryotes and eukaryotes, along with the role of various types of RNA in gene expression.

Tasks:

1. Compare the genome organization in prokaryotes and eukaryotes, emphasizing the differences in regulatory genetic elements.

2. Discuss the role of various types of RNA (rRNA, mRNA, tRNA, U-family RNA, 7S-RNA) in the process of gene expression.

3. Analyze how these RNA molecules contribute to transcription, translation, and post-transcriptional regulation in gene expression.

Keywords: Genome organization, gene regulation, prokaryotes, eukaryotes, rRNA, mRNA, tRNA, U-family RNA, 7S-RNA, transcription, translation, post-transcriptional regulation

What is genome?

✓ **A GENOME IS ALL THE GENETIC INFORMATION OF AN ORGANISM**
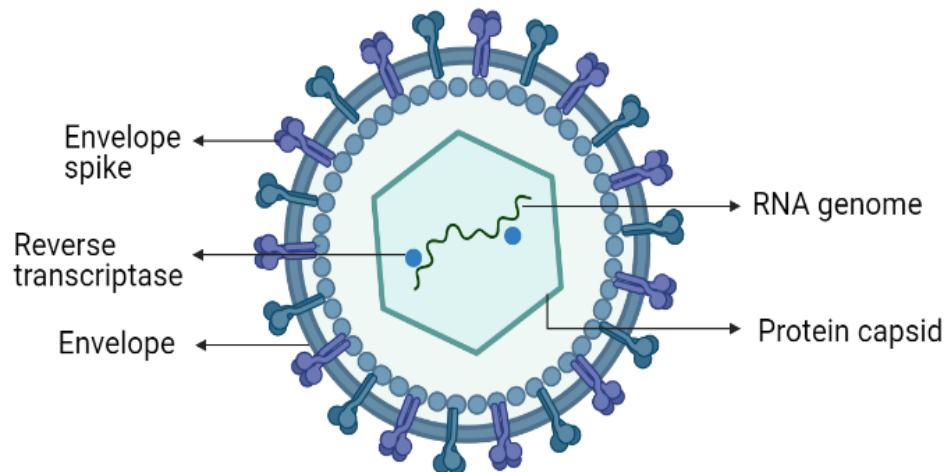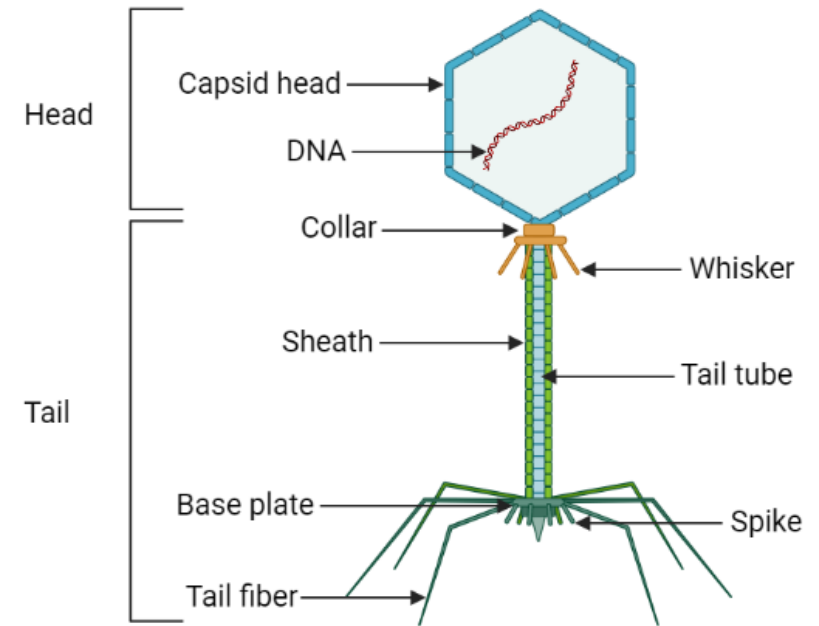
- It provides all of the information the organism requires to **function**.
- In living organisms, the genome is stored in long molecules of DNA called **chromosomes**.
- In eukaryotes, each cell's genome is contained within a membrane-bound structure called the **nucleus**.
- Prokaryotes, which contain no inner membranes, store their genome in a region of the cytoplasm called the **nucleoid**.
- The study and analysis of genomes is called **genomics**.

Note: There are 23 pairs of chromosomes in the human genome. Between 1990 and 2003, all twenty-three pairs were fully sequenced through an international research undertaking known as the Human Genome Project.
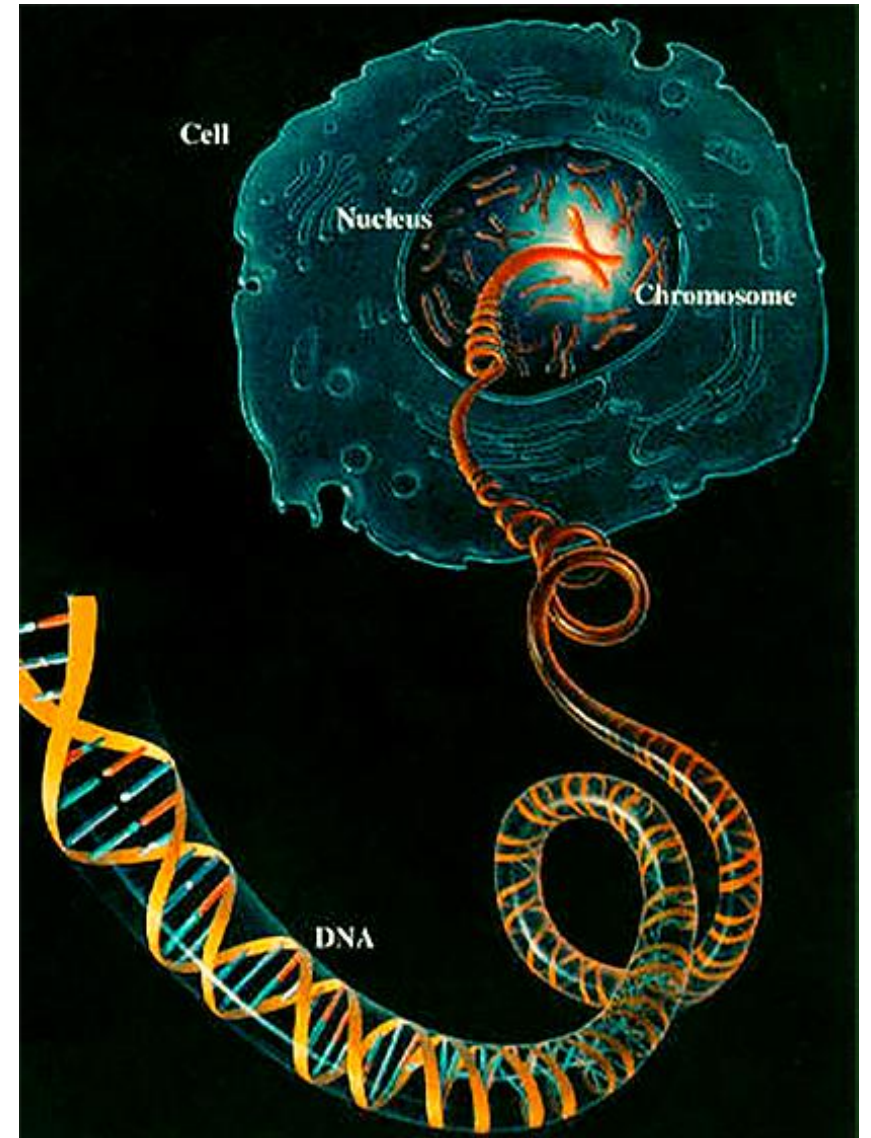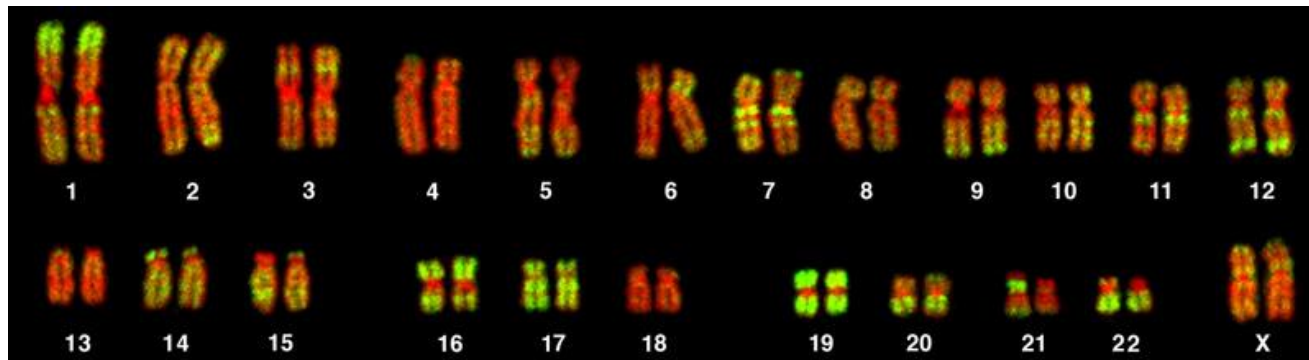
Organization of genomes

✓The term "genome" was proposed by G. Winkler in 1920 to describe the totality of genes contained in the haploid set of chromosomes of organisms of the same biological species.

✓The genome includes the entire set of DNA molecules of the cell (in the case of a number of viruses, they speak of genomic RNA).

Head — Capsid head, DNA
Collar — Whisker
Sheath — Tail tube
Base plate — Spike
Tail fiber
Tail

Envelope spike
Reverse transcriptase
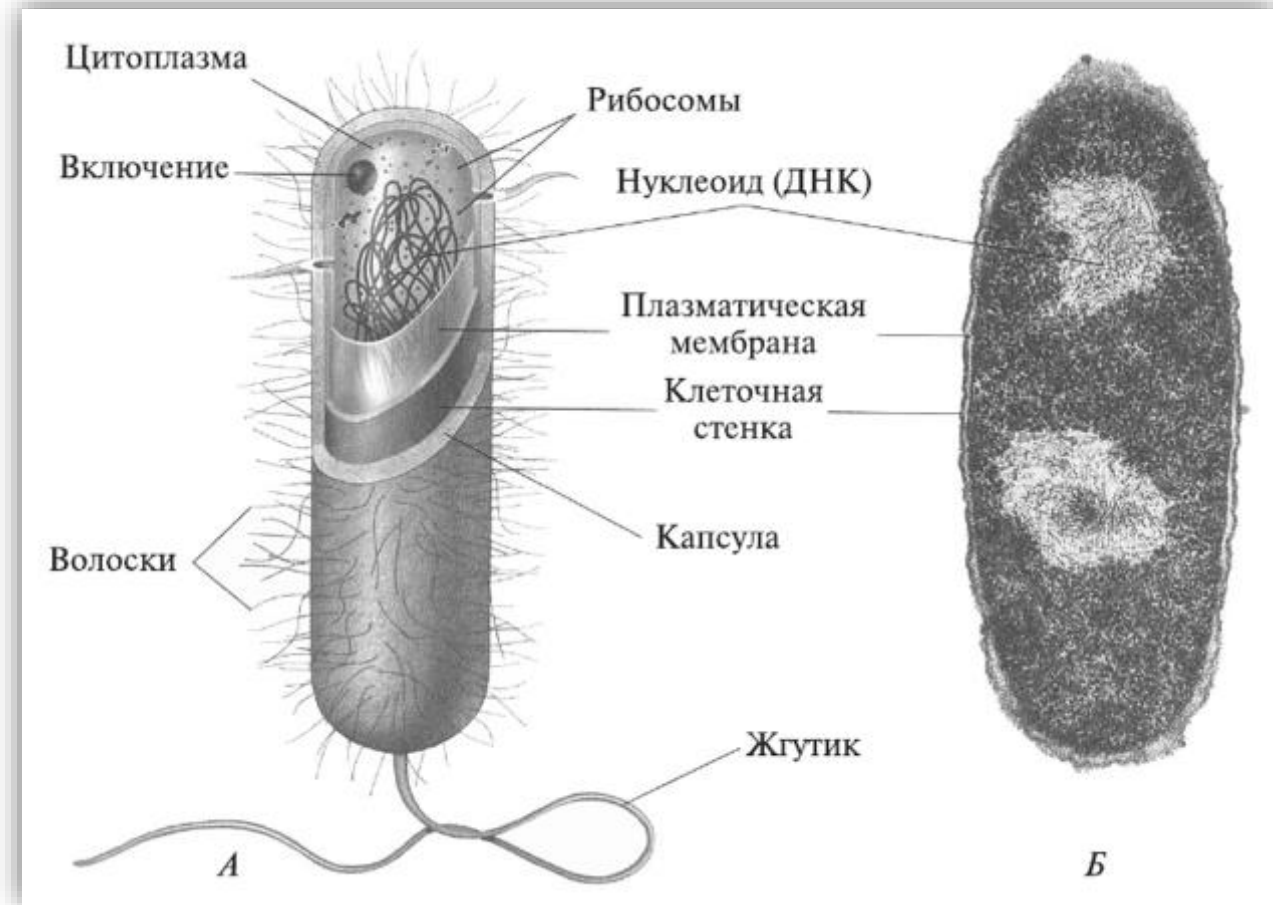Envelope
RNA genome
Protein capsid

- **Features:**

1. *Smaller size*

2. *Diverse genome types* (DNA or RNA; Single-stranded (ss) or double-stranded (ds); Positive-sense (+) or negative-sense (−) RNA)

3. *different physical forms* (Linear or circular; Segmented or non-segmented)

4. *Compactness and gene density* (Overlapping genes; Polyprotein strategy; Alternative splicing)

5. *High mutation rates* in RNA viruses

6. *Lower mutation rates* in DNA viruses

7. *Recombination and reassortment*

Prokaryotic genome

- **Features:**

1. *Small genome size (0.5–10 Mb).*
2. *Circular DNA is most common, often with plasmids.*
3. *Haploid, with a single copy of each gene.*
4. *High gene density, with few non-coding sequences.*
5. *Operon structure for coordinated gene expression.*
6. *Horizontal gene transfer is a key factor in evolution.*
7. *Rapid replication with a single origin of replication.*
8. *No nucleus; the DNA is found in the nucleoid.*
9. *Mobile genetic elements such as transposons are common.*

In 1956, F. Jacob and E. Volman proposed a ring model for the organization of bacterial chromosomes.

Цитоплазма
Рибосомы
Включение
Нуклеоид (ДНК)
Плазматическая мембрана
Клеточная стенка
Капсула
Волоски
Жгутик
А
Б

The genome size is 4.6 million bp.

Most of the **prokaryotic genome** (usually 80-90%) is made up of sequences encoding proteins and RNA
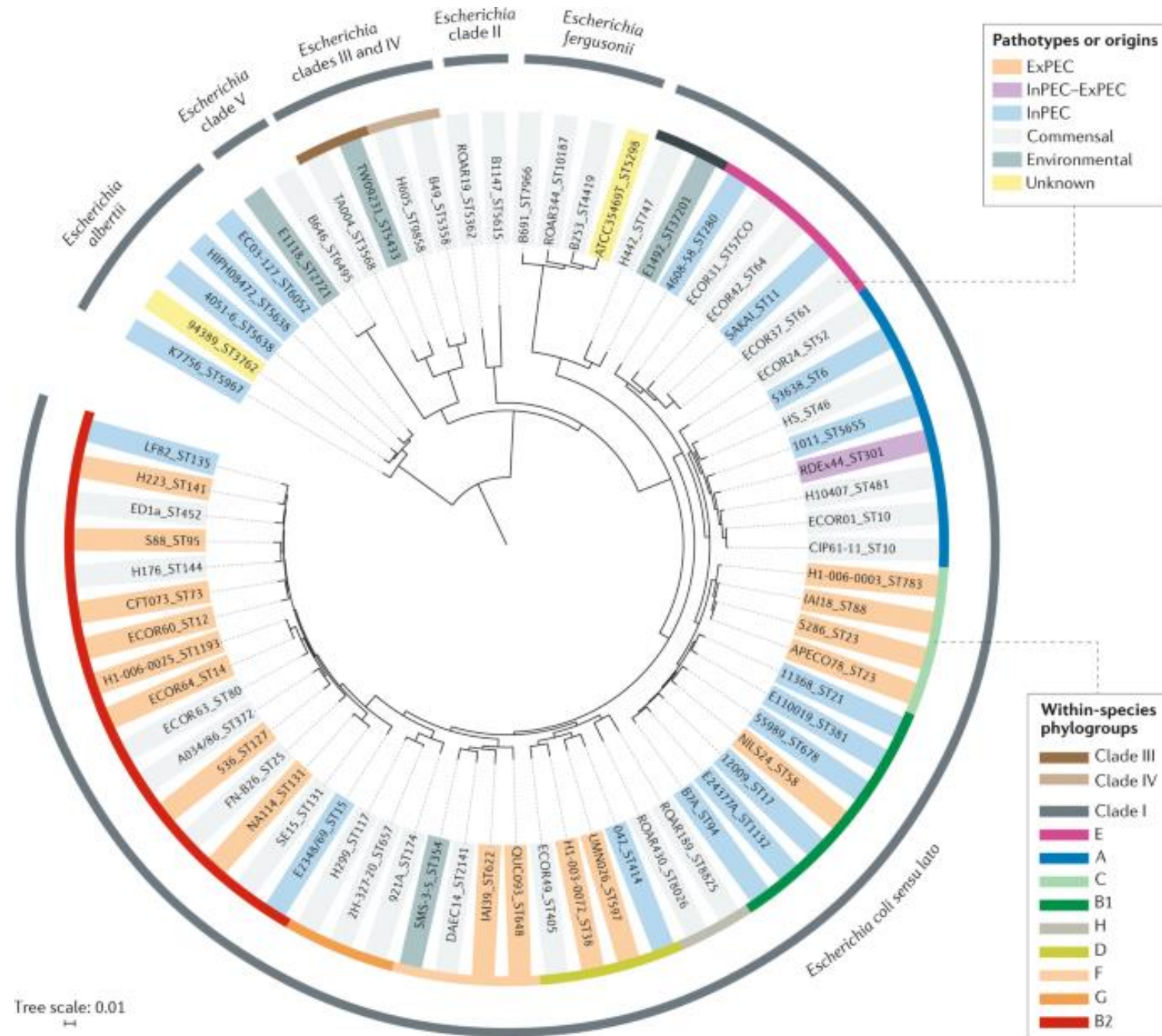
Organisms	Source	Genome size (Mb)	Total genes [1]
Phage/viruses, organelle			
1 mitochondrion	Human mitochondrion	16.6×10^3	40
2 SV40	Simian virus 40	5.2×10^3	8
3 TMV	Tobacco Mosaic Virus (RNA)	6.4×10^3	4
4 HIV	AIDS virus (RNA)	9.3×10^3	10
5 adenovirus 2	Human adenovirus	35.9×10^3	11
6 λ phage	lysogenic phage	48.5×10^3	50
7 T4 phage	DNA virus E. coli	169×10^3	300
Bacteria			
8 Mycoplasma genitalium		580×10^3	470
9 Rickettsia prowazeki		1.11×10^6	834
10 Streptococcus pyogenes		1.85×10^6	1752
11 Bacillus subtilis		4.21×10^6	4100
12 Esherichia coli		4.64×10^6	4288
13 Mycobacterium tuberculesis		4.41×10^6	3924
14 Helicobacter pylori		1.668×10^6	1590 [1]
		1.644×10^6	1495 [1]
Eucaryotes			
15 Saccharomyces cerevisiae	budding yeast	12×10^6	5885
16 Caenorhabditis elegans	worm, Nematoda	97×10^6	19100
17 Arabidopsis thaliana	plant, Angiospermae	120×10^6	25500
18 Drosophila melanogaster	fruit fly, Diptera	165×10^6	13000
19 Mus musculus	mouse	2.5×10^9	30000
20 Homo sapiens	humans	3.2×10^9	31000

[1] Two strains isolated from two geographically separated regions of USA (32).

Eukaryotic genome
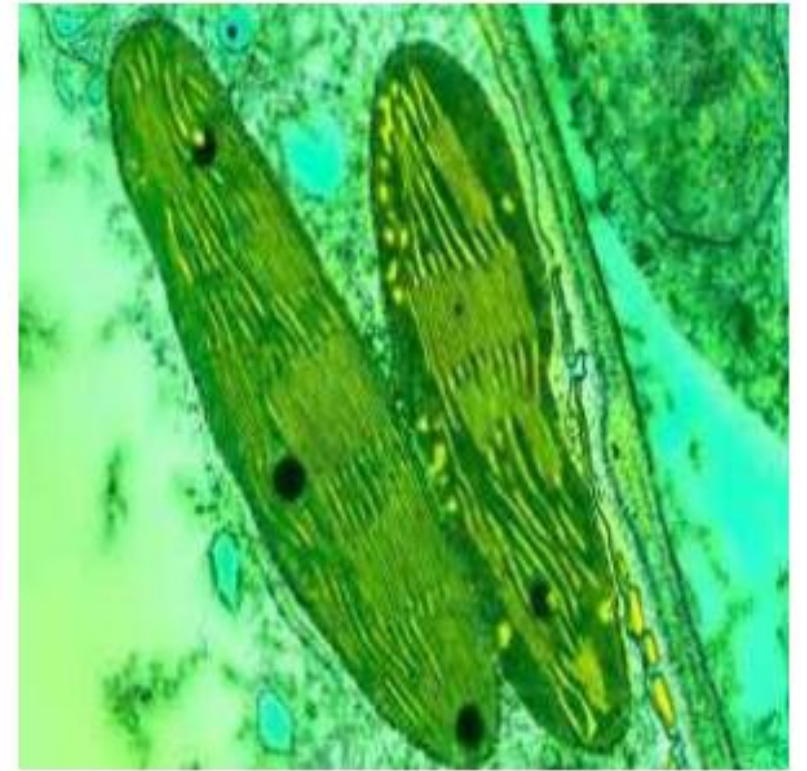
- The **eukaryotic genome** is much more complex than that of prokaryotes.
- The **genetic apparatus** of a eukaryotic cell is isolated in the form of a cell nucleus, inside which the main carriers of heredity, the chromosomes, are located.
- The **number of chromosomes** is species-specific and ranges from two (horse roundworm) to a thousand (lower plants).

Organism	Chromosomes	Estimated genome size	Estimated gene number	Estimated gene density gene/ base pairs
Homo Sapiens (human)	46	290×10^6	30 000	1/ 100 000
Mus musculus (mouse)	40	250×10^6	30 000	1/ 100 000
Drosophila melanogaster (fruit fly)	8	180×10^6	13 600	1/ 9000
Arabidopsis thaliana (plant)	5	125×10^6	25 500	1/ 4000
Caenorhabditis elegans (roundworm)	6	97×10^6	19 100	1/ 5000
Saccharomyces cerevisiae (yeast)	16	12×10^6	6 300	1/ 2000
Escherichia coli (bacteria)	1	4.7×10^6	3 200	1/ 1400
Haemophilus influenzae (bacteria)	1	1.8×10^6	1 700	1/ 1000

2.5.1 Human-mouse genome sequence comparisons

The C-paradox is the lack of correlation between the physical size of the genome and the complexity of organisms.

Reasons for redundancy:
1. Large size of genes (due to the presence of introns).
2. Presence of repeated sequences. Both genes and non-coding regions are repeated.
3. The presence of a large number of non-coding sequences

Disadvantages of "excessive" DNA: - increase in DNA synthesis time; - it is more difficult to organize DNA duplication; - high energy intensity - for 1 nucleotide to be included in the DNA chain, you need to spend ~ 60 ATP molecules.
Indeterminate consequence: - due to the dependence of the size of the nucleus on the amount of DNA, an increase in the size of the cell occurs.
Pluses of "excessive" DNA: - it becomes possible to create a complex regulatory apparatus that allows you to raise the body to a higher evolutionary level.

Organization of genomes:

❖ **Genes**:
  - Genes are segments of DNA that carry the information for synthesizing RNA and proteins. Proteins are key components responsible for carrying out cellular functions.
  - Genes contain coding sequences (**exons**) and non-coding sequences (**introns**), which are removed during RNA processing.

❖ **Regulatory Elements**:

❖ **Promoters** and **enhancers** are DNA regions that regulate gene activity. Promoters determine where and when a gene will be active, while enhancers boost the transcription of genes.

❖ **Non-coding Sequences**:
  These regions of DNA do not code for proteins but may play important regulatory and structural roles. For example, they can include sequences that regulate gene expression or regions responsible for maintaining chromosome structure.
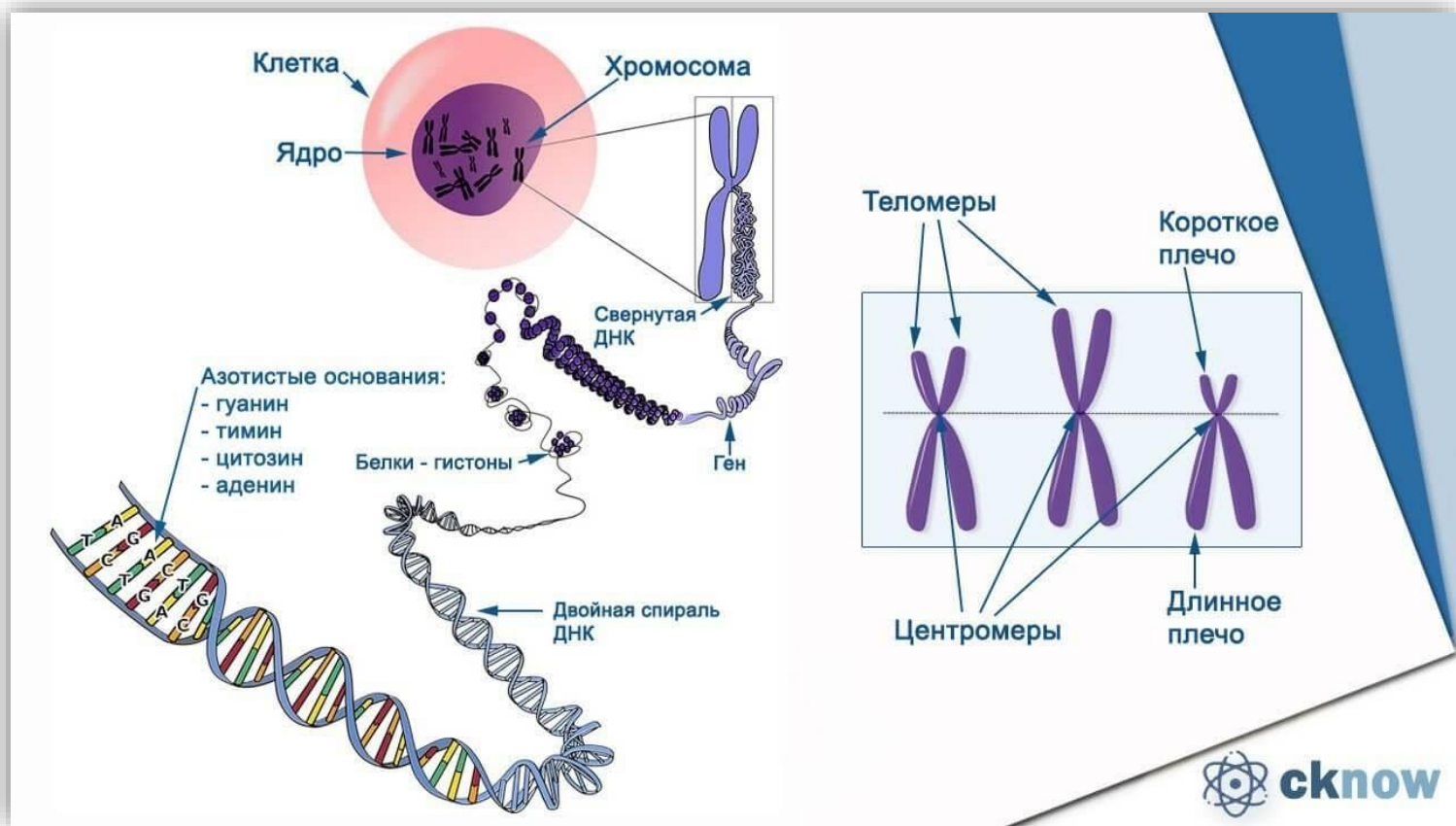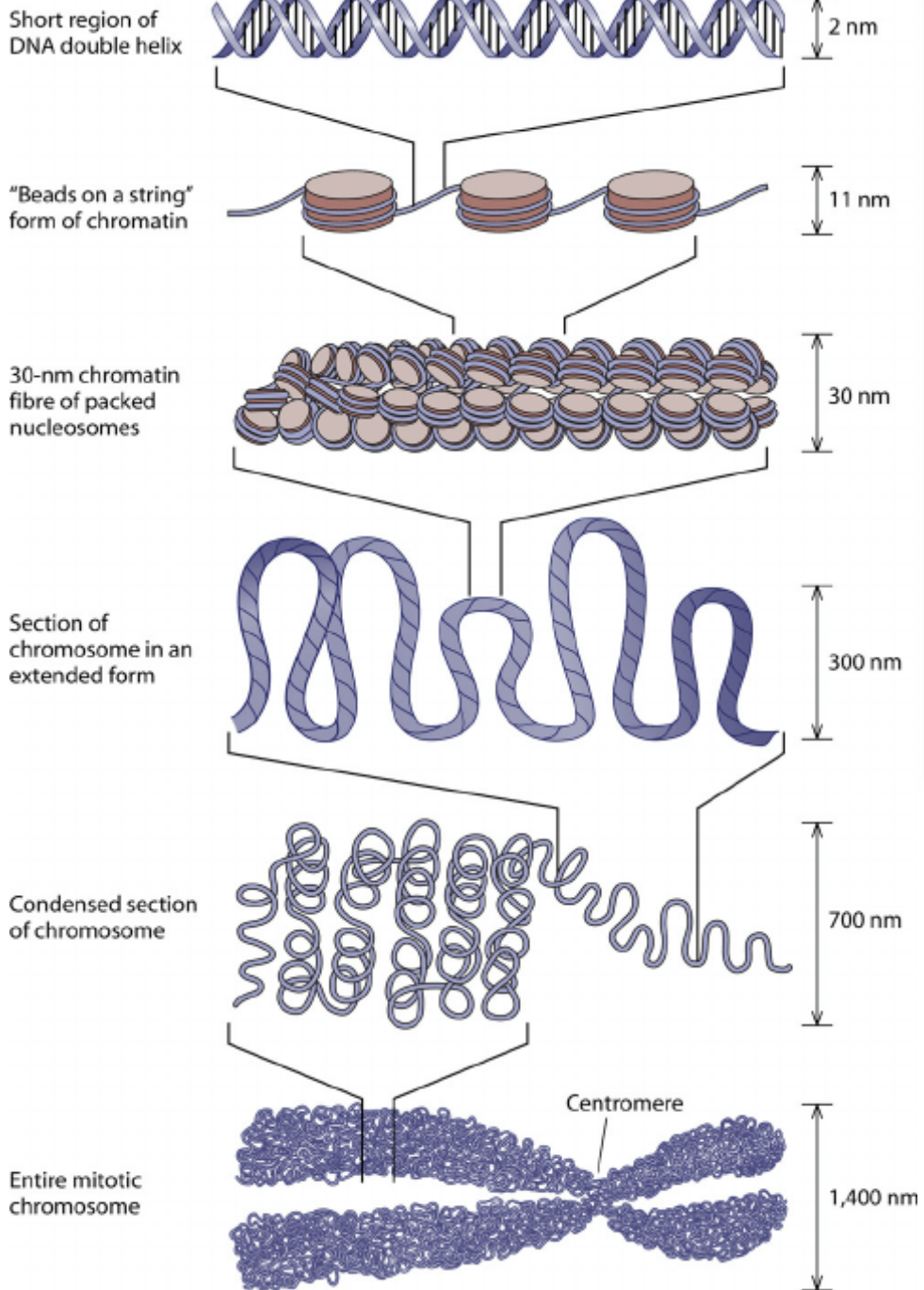
❖ **Chromosomes**:
  DNA is packaged in the cell nucleus as chromosomes, which allow it to take up less space and facilitate the distribution of genetic material during cell division. In eukaryotes (such as humans), each chromosome is made up of a linear DNA molecule associated with proteins like histones.
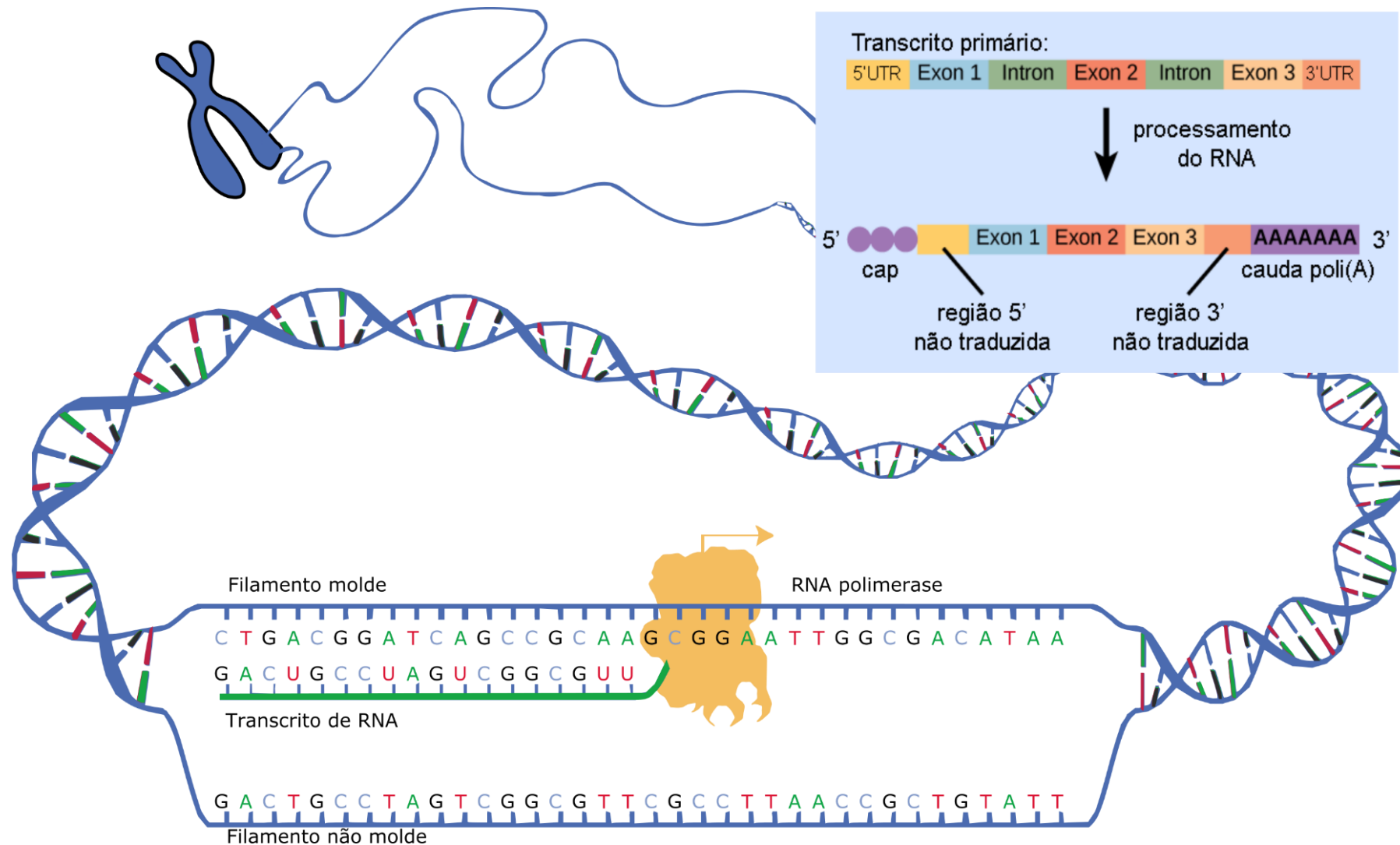
❖ **Genomes in Different Organisms**:
  In prokaryotes (bacteria and archaea), genomes typically take the form of a circular DNA molecule found in the cytoplasm, while in eukaryotes, genomes are usually linear and stored in the nucleus.
  The size of genomes can vary significantly, from a few thousand to billions of base pairs, depending on the organism.
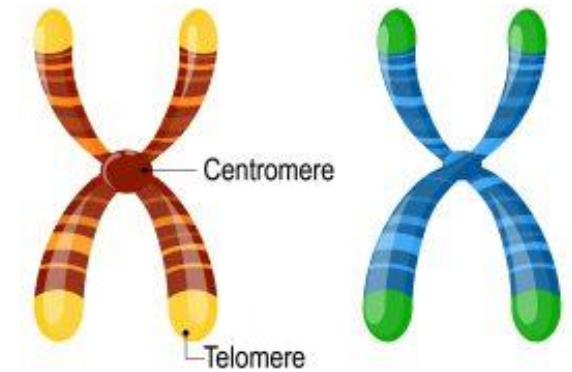
Short region of DNA double helix — 2 nm

"Beads on a string" form of chromatin — 11 nm

30-nm chromatin fibre of packed nucleosomes — 30 nm

Section of chromosome in an extended form — 300 nm

Condensed section of chromosome — 700 nm

Entire mitotic chromosome — Centromere — 1,400 nm

Клетка

Ядро

Хромосома

Свернутая ДНК

Азотистые основания:
- гуанин
- тимин
- цитозин
- аденин

Белки - гистоны

Ген

Двойная спираль ДНК

Теломеры

Короткое плечо

Центромеры

Длинное плечо

cknow

Individual genetic elements with a strictly **specific nucleotide sequence** encoding certain products are **genes**. Some of them code for **proteins**, others only for **RNA** molecules.

Transcrito primário:

| 5'UTR | Exon 1 | Intron | Exon 2 | Intron | Exon 3 | 3'UTR |

processamento do RNA

5' ●●● Exon 1 Exon 2 Exon 3 AAAAAAA 3'

cap

região 5' não traduzida

cauda poli(A)

região 3' não traduzida

Filamento molde

RNA polimerase

C T G A C G G A T C A G C C G C A A G C G G A A T T G G C G A C A T A A

G A C U G C C U A G U C G G C G U U

Transcrito de RNA

G A C T G C C T A G T C G G C G T T C G C C T T A A C C G C T G T A T T

Filamento não molde

- The content of **satellite DNA** in the eukaryotic genome is from **5 to 50%** of the total DNA.
- This DNA is predominantly found in the centromeric and telomeric regions of chromosomes, where it performs structural functions.
- Satellite DNA consists of tandem repeats from 1 to 20 or more bp in length. Due to the simplicity of organization and numerous copies, this DNA has the ability to rapidly renature.

CHROMOSOME

Centromere

Telomere

5'–CACACACACACAC–3'
3'–GTGTGTGTGTGTG–5'

5'–AAAAAAAAAAAAAAA–3'
3'–TTTTTTTTTTTTTTT–5'

Tandem Repeats

≡ PEDIAA

8 Repeats

Genome 1   CTAGAGATAGATAGATAGATAGATAGATAGATAGATAGATACTAGACTAGACTA
Genome 2   CTAGAGATAGATAGATAGATAGATAGATAGATAGATAGATACTAGACTAGACTA
Genome 3   CTAGAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATACTAGACTAGACTA

9 Repeats

Genetic information moves from DNA to RNA to protein
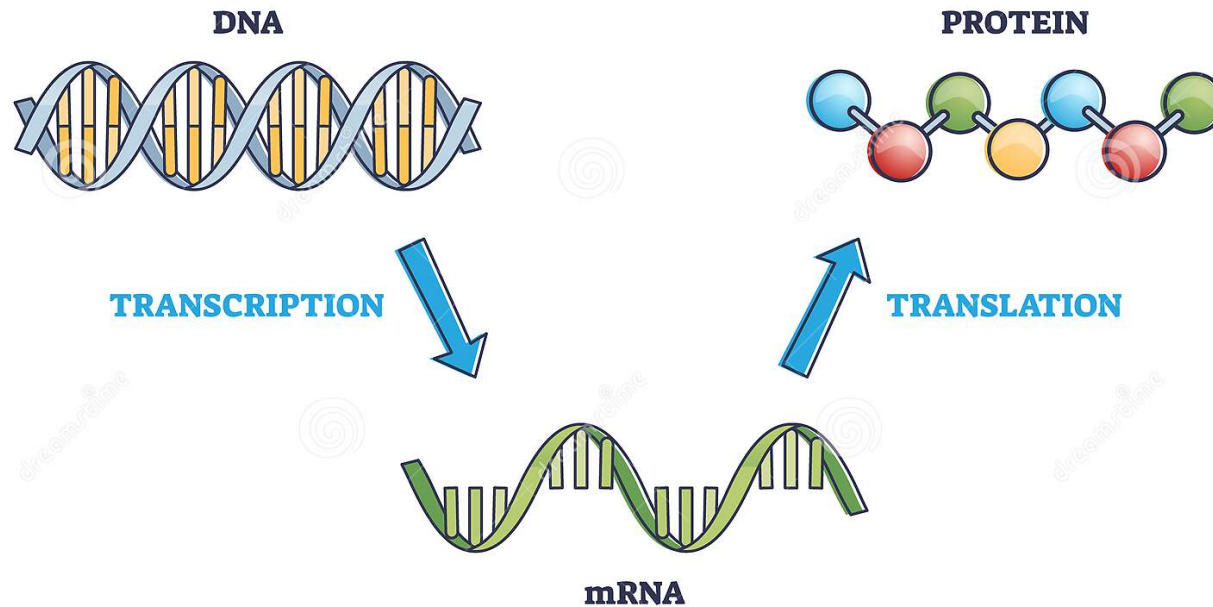
Gene expression is the process by which DNA directs the synthesis of functional products such as proteins. Cells can regulate gene expression at various stages. This allows organisms to generate different types of cells and allows cells to adapt to internal and external factors.
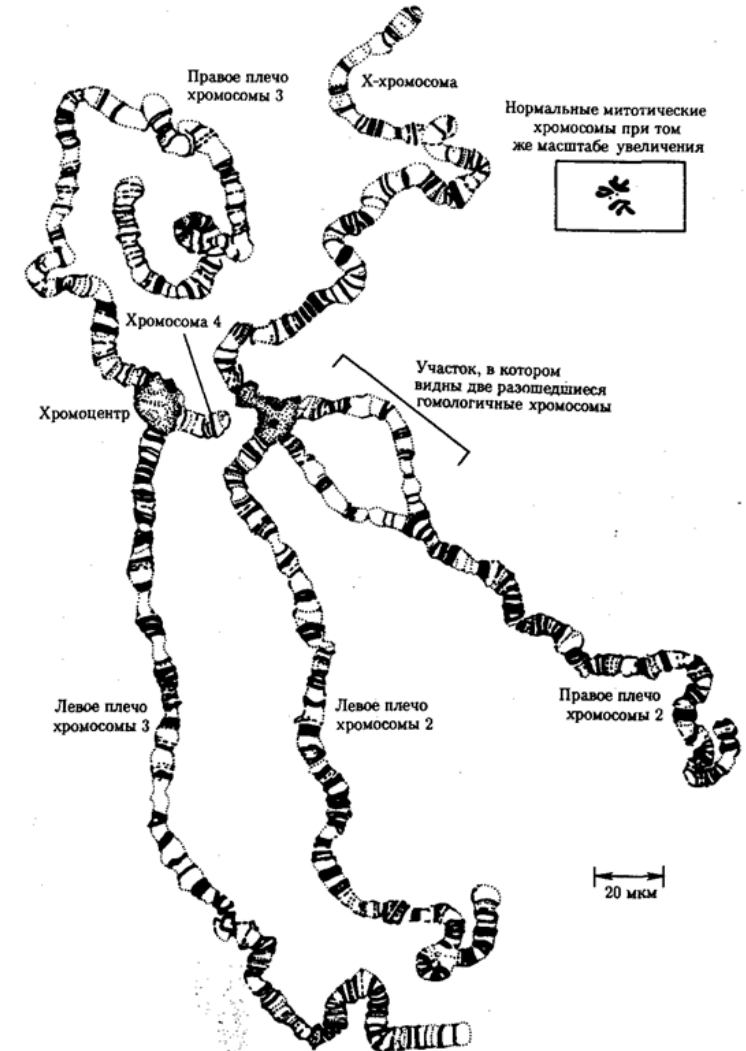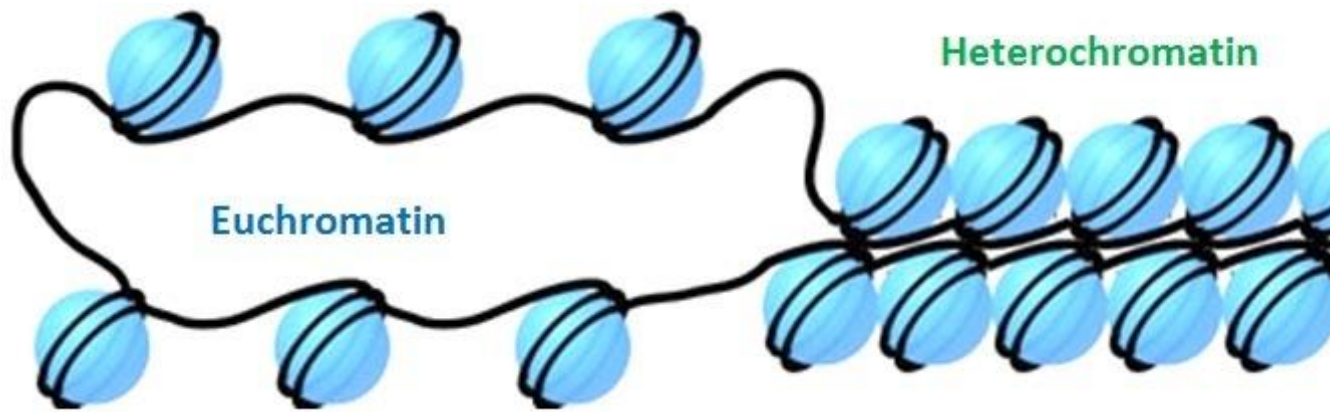
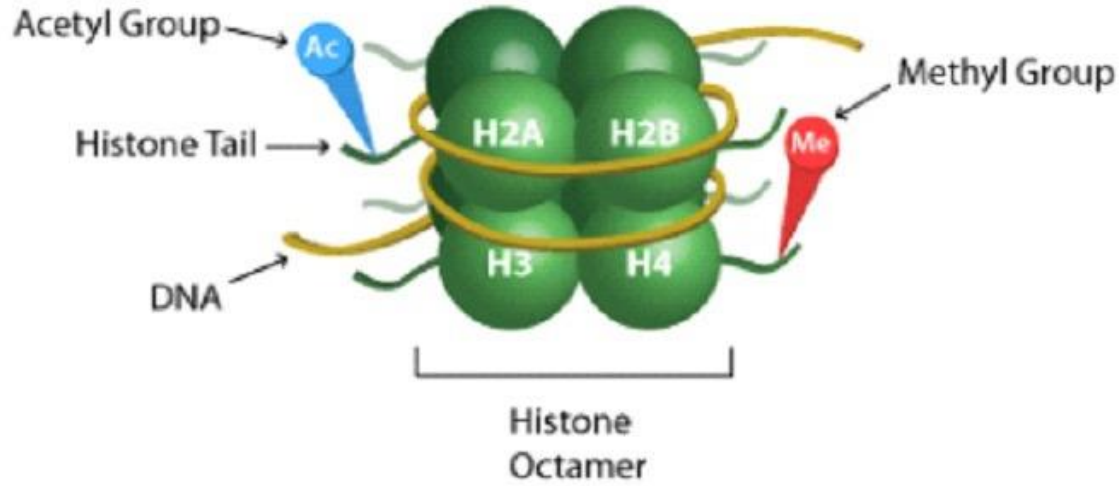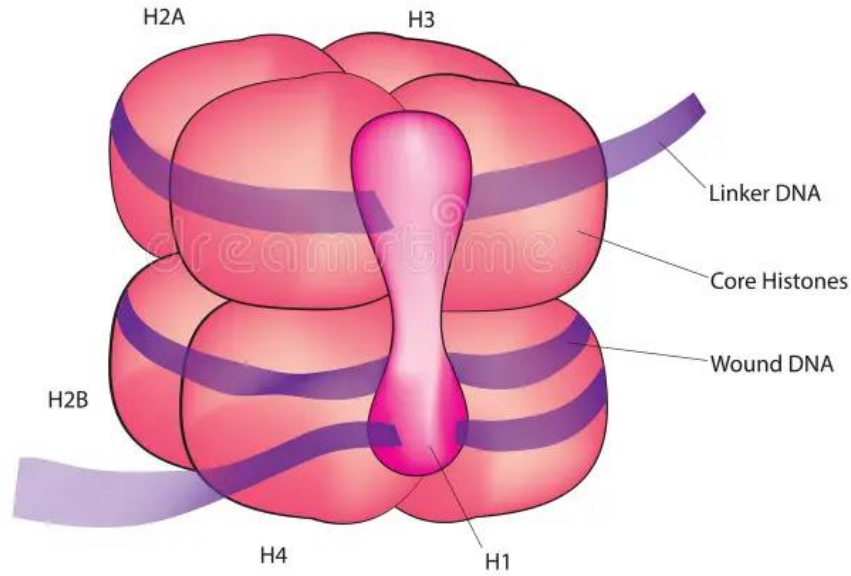# Types of regulation of eukaryotic gene expression
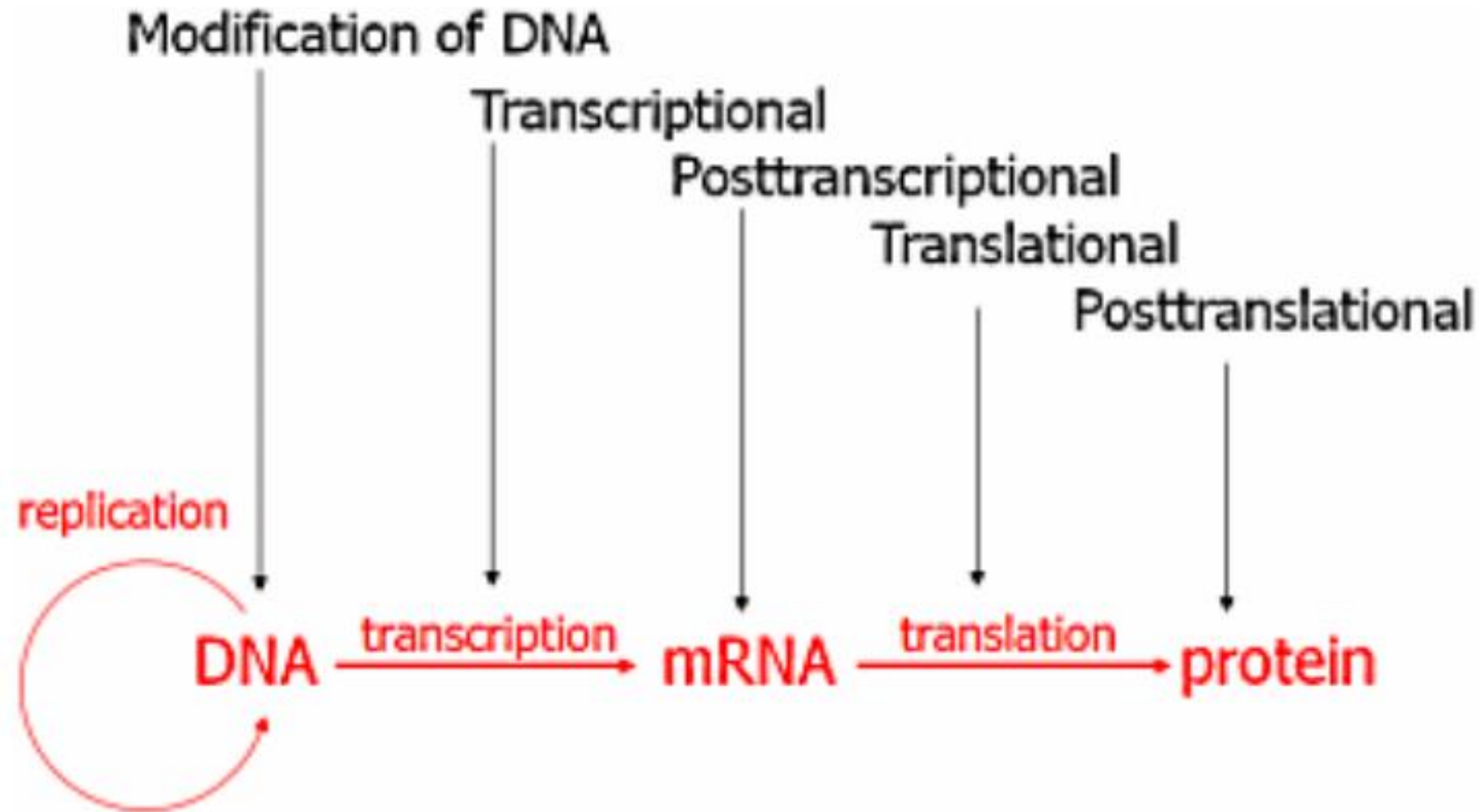
- **DNA methylation and demethylation.**
- DNA methylation occurs in the regulatory regions of the gene. Cytosine is methylated in the CG sequence, after which the gene is inactivated.
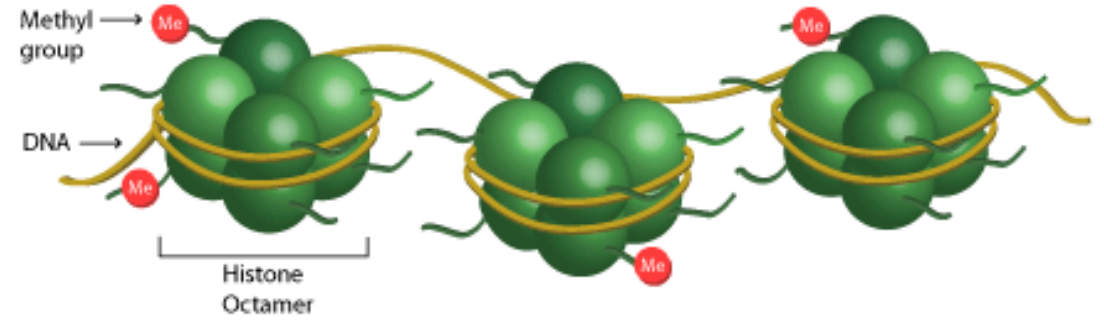- Demethylation restores gene activity. The process is regulated by the enzyme methyltransferase.
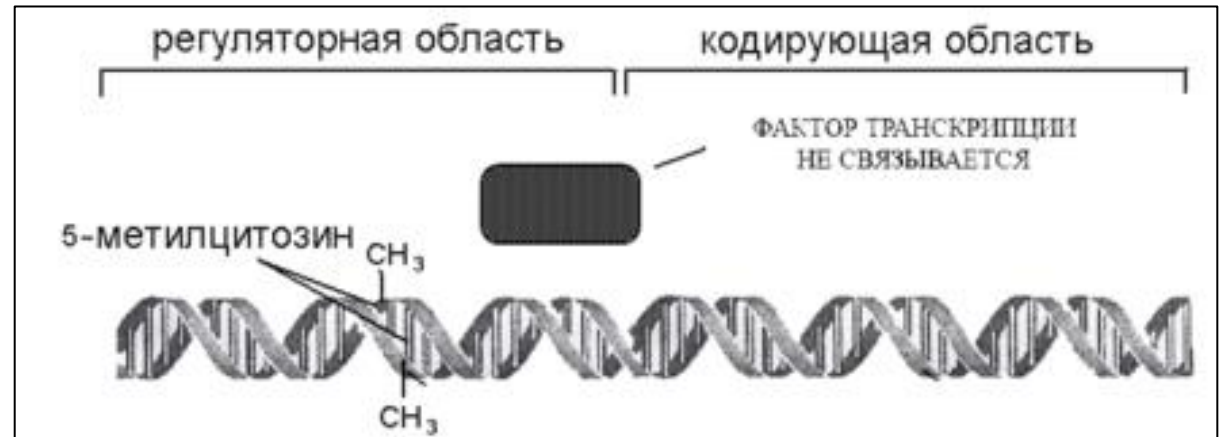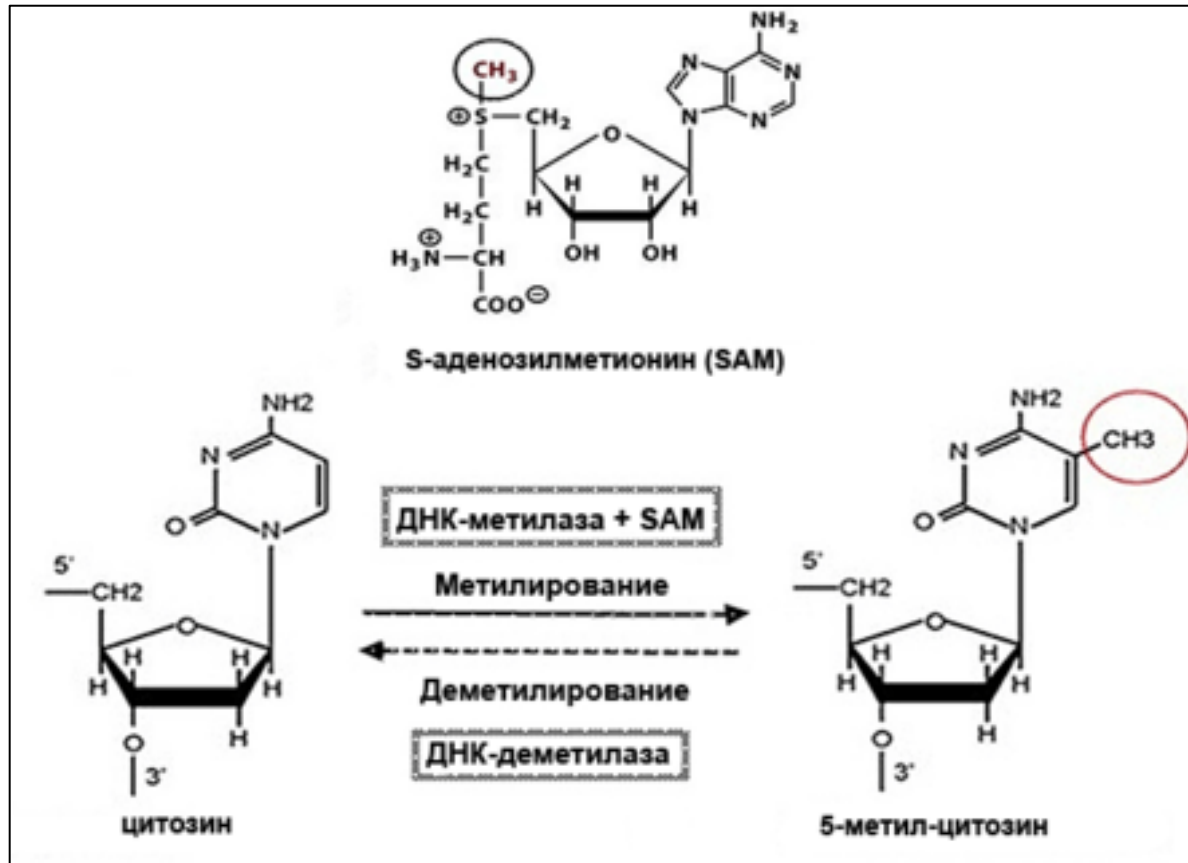
- **Condensation and decondensation of chromatin**.
- This is the most versatile method for regulating transcription. When certain genes need to be expressed, the chromatin decondenses at that location.
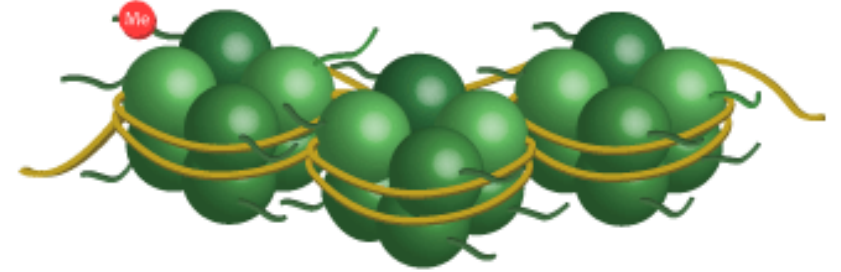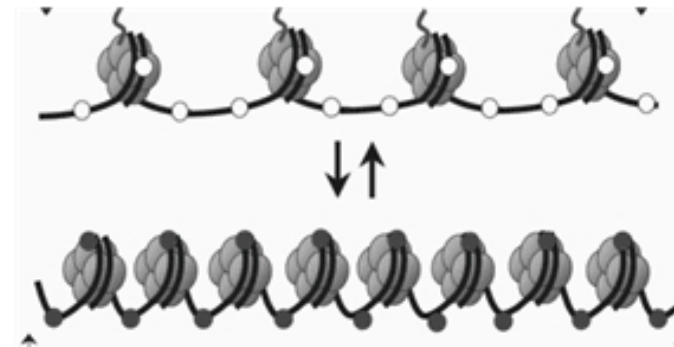

Heterochromatin

Euchromatin

Правое плечо хромосомы 3

X-хромосома

Нормальные митотические хромосомы при том же масштабе увеличения

Хромосома 4

Участок, в котором видны две разошедшиеся гомологичные хромосомы

Хромоцентр

Левое плечо хромосомы 3

Левое плечо хромосомы 2

Правое плечо хромосомы 2

20 мкм

NUCLEOSOME

H2A
H3
Linker DNA
Core Histones
Wound DNA
H2B
H4
H1

Euchromatin

Methyl group → Me
DNA →
Histone Octamer

Heterochromatin

Me

Acetyl Group → Ac
Histone Tail →
Methyl Group → Me
H2A H2B
H3 H4
DNA →
Histone Octamer

нуклеосома

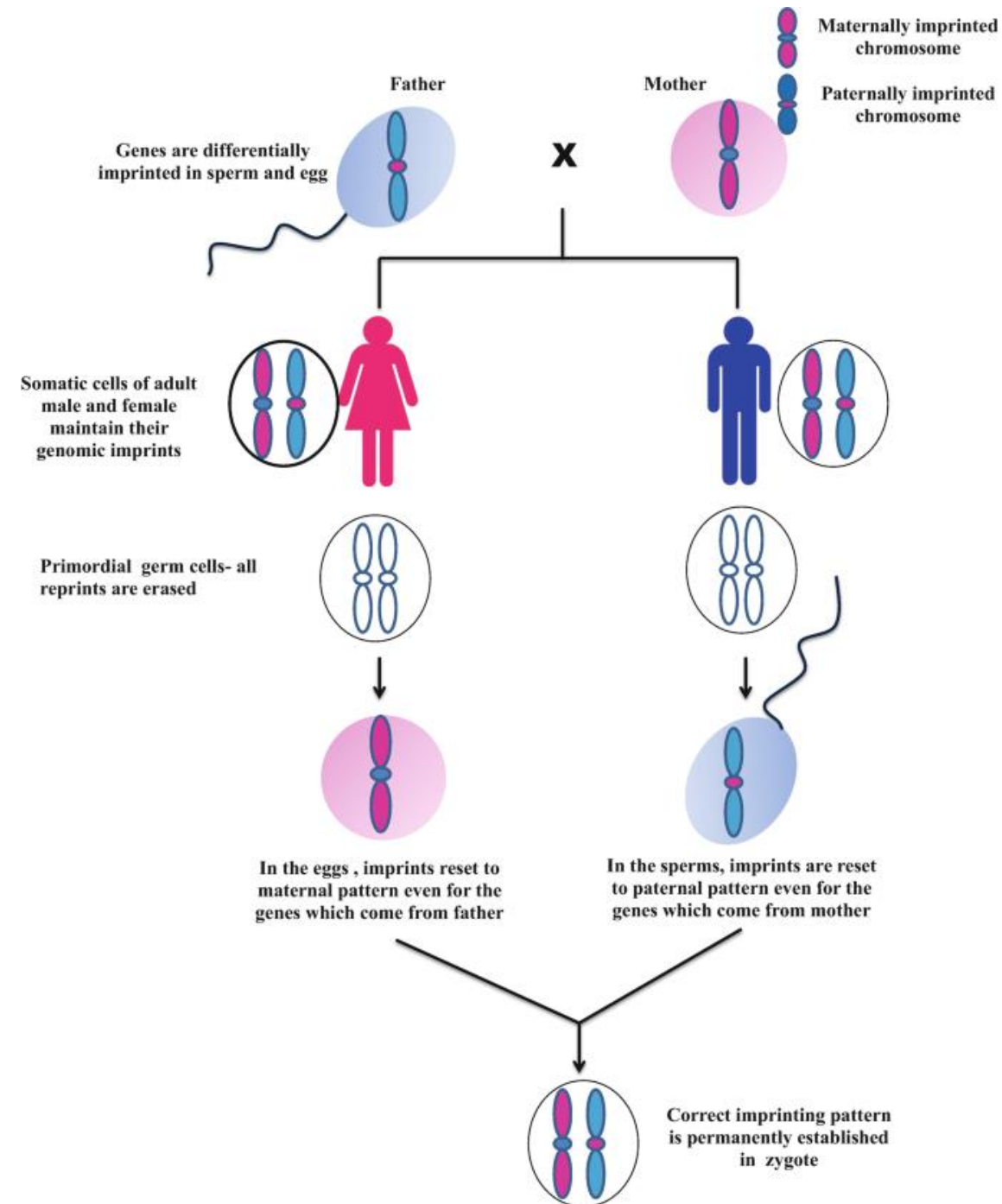ДНК деметилирована, гистоны ацетилированы, транскрипция возможна

↓↑

ДНК метилирована, гистоны деацетилированы, транскрипция невозможна (черные точки – метилцитозин)
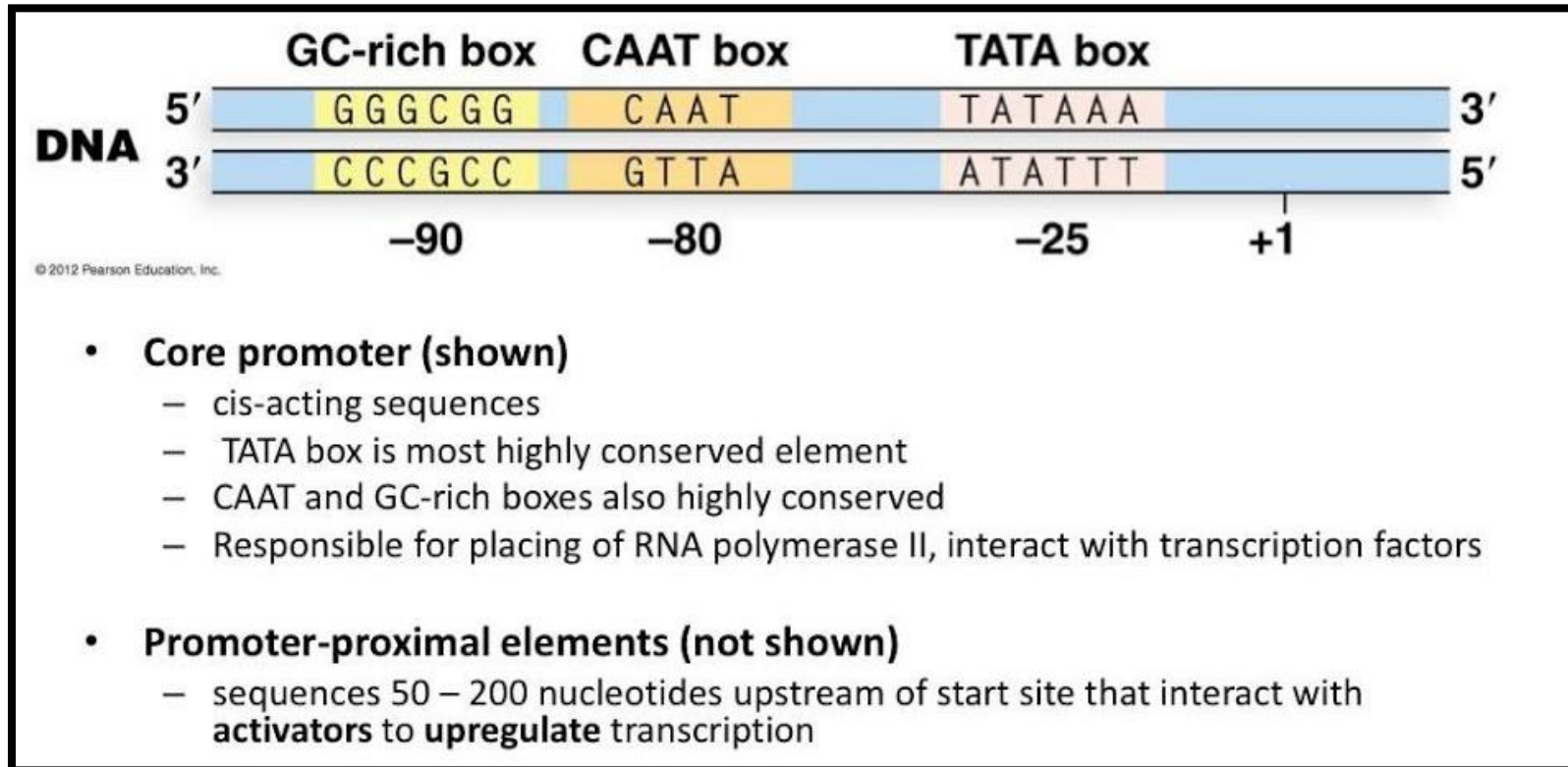
- **Hormonal regulation.**
- With hormonal regulation, genes are activated in response to an external chemical signal (the entry of a certain hormone into the cell). This hormone starts those genes that have specific nucleotide sequences in the regulatory regions.
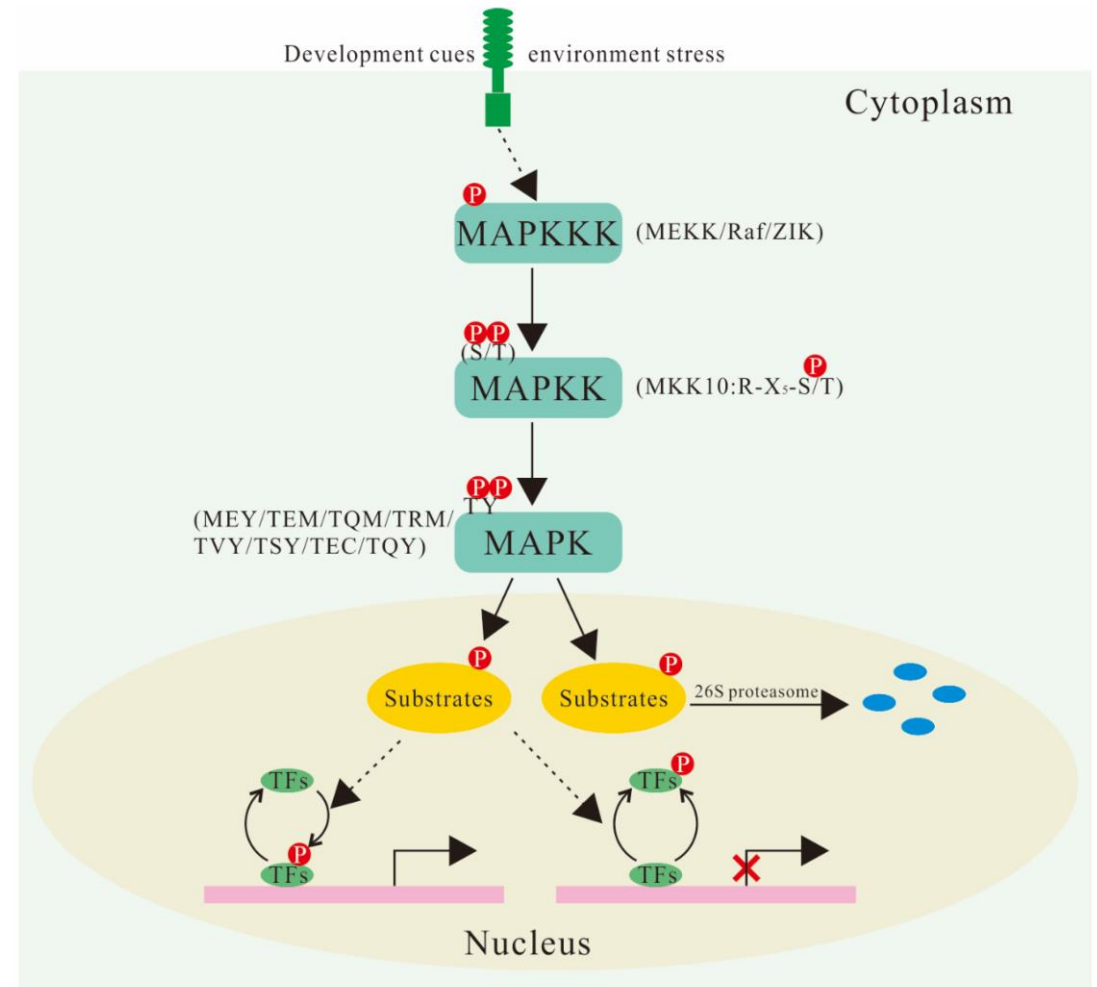
- **Genomic imprinting**.
- This is a poorly understood way of regulating gene expression in eukaryotes. It is possible only in diploid organisms and is expressed in the fact that the activity of genes depends on which parent they were obtained from. Turning off genes is carried out by DNA methylation.

Maternally imprinted chromosome
Paternally imprinted chromosome

Father

Mother

X

Genes are differentially imprinted in sperm and egg

Somatic cells of adult male and female maintain their genomic imprints

Primordial germ cells- all reprints are erased

In the eggs, imprints reset to maternal pattern even for the genes which come from father

In the sperms, imprints are reset to paternal pattern even for the genes which come from mother

Correct imprinting pattern is permanently established in zygote

Despite the individuality of the set of regulatory elements in the structural genes of eukaryotes, each of them has a promoter region (TATA box, or Hogness box) of eight nucleotides, including the TATA sequence; CCAAT sequence (CAT box); a section of repeated GC dinucleotides (GC-box). These elements are located at a distance of 25, 75 and 90 bp. from the site of initiation, respectively.
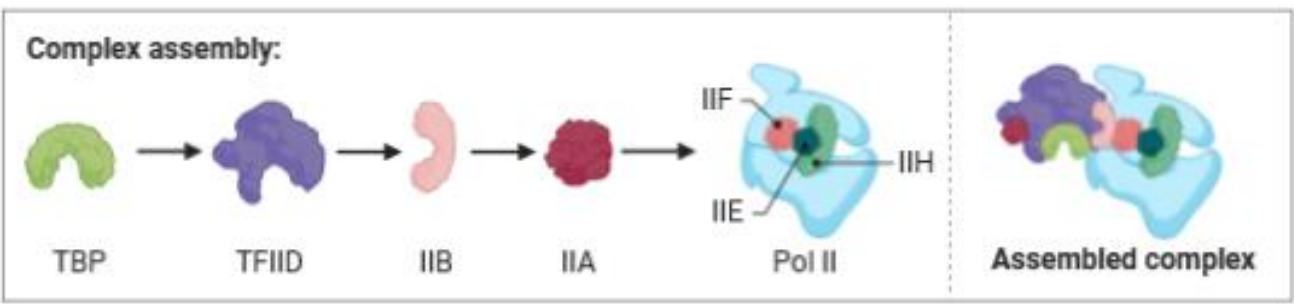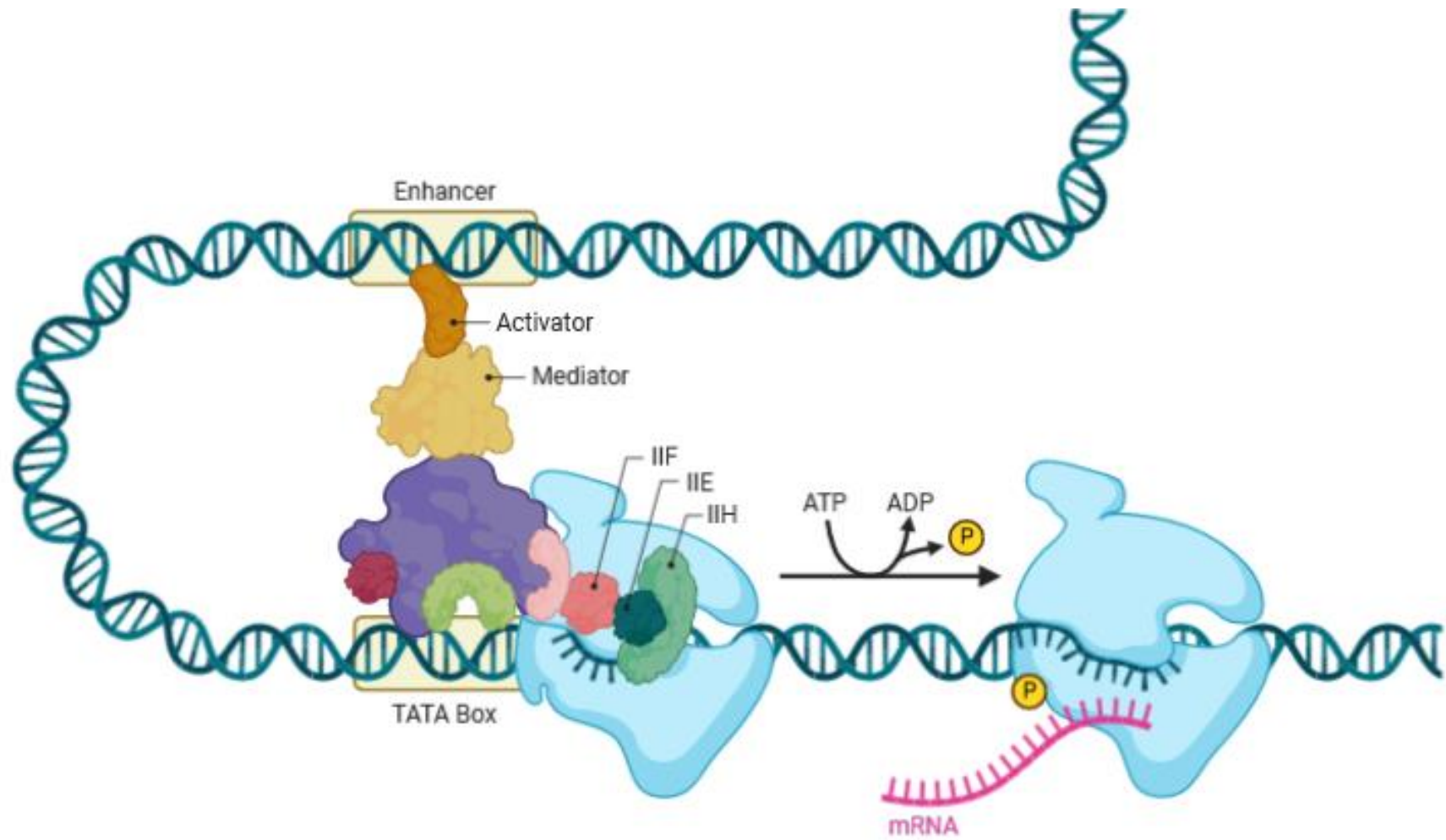

© 2012 Pearson Education, Inc.

- **Core promoter (shown)**
  - cis-acting sequences
  - TATA box is most highly conserved element
  - CAAT and GC-rich boxes also highly conserved
  - Responsible for placing of RNA polymerase II, interact with transcription factors

- **Promoter-proximal elements (not shown)**
  - sequences 50 – 200 nucleotides upstream of start site that interact with **activators** to **upregulate** transcription
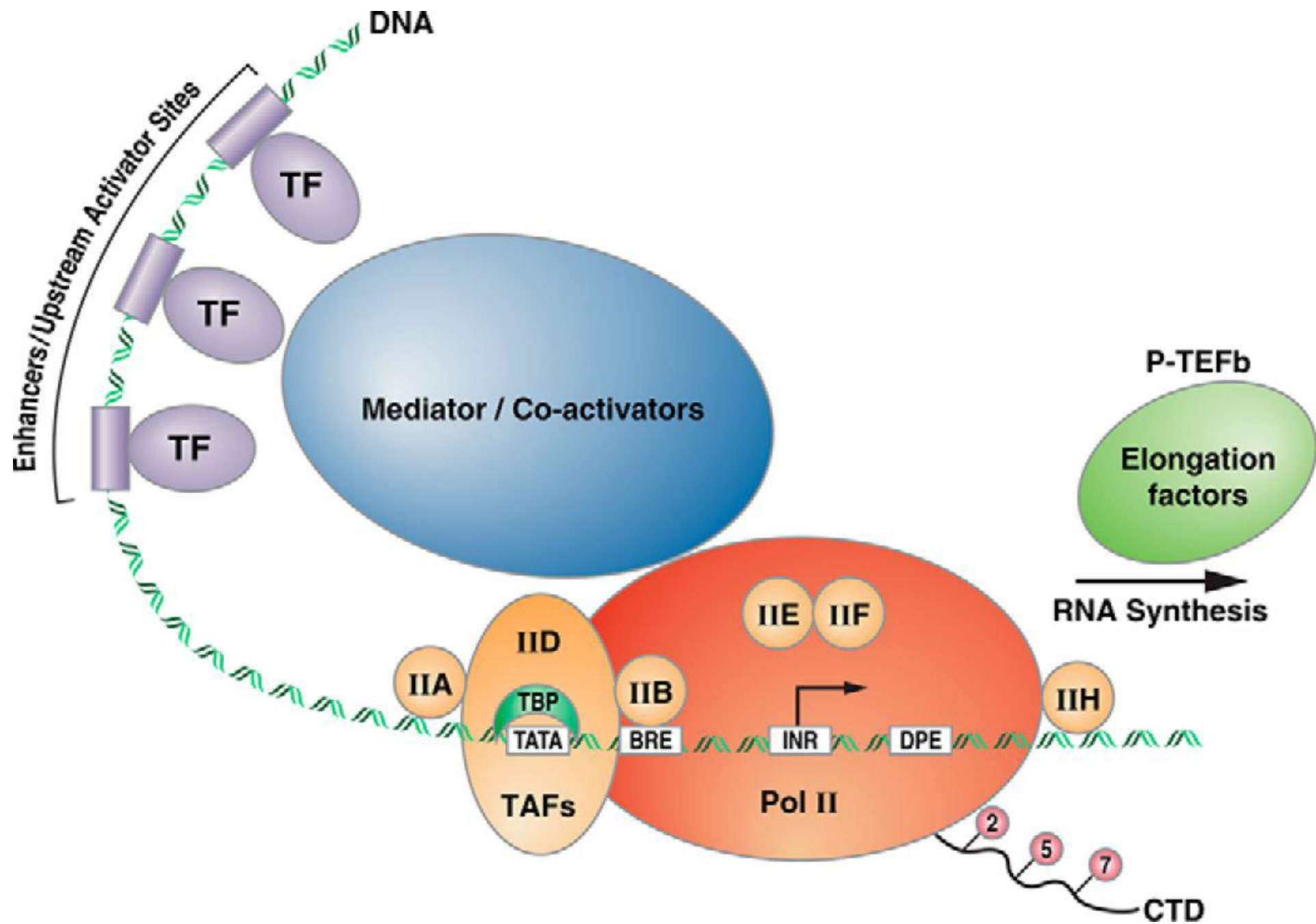
Unlike prokaryotes, eukaryotes mostly lack operons, i.e., each eukaryotic structural gene has its own set of regulatory elements. A significant role in the regulation of transcription in eukaryotes, in addition to being mediated by the interaction between DNA and proteins, is also played by protein-protein interactions.
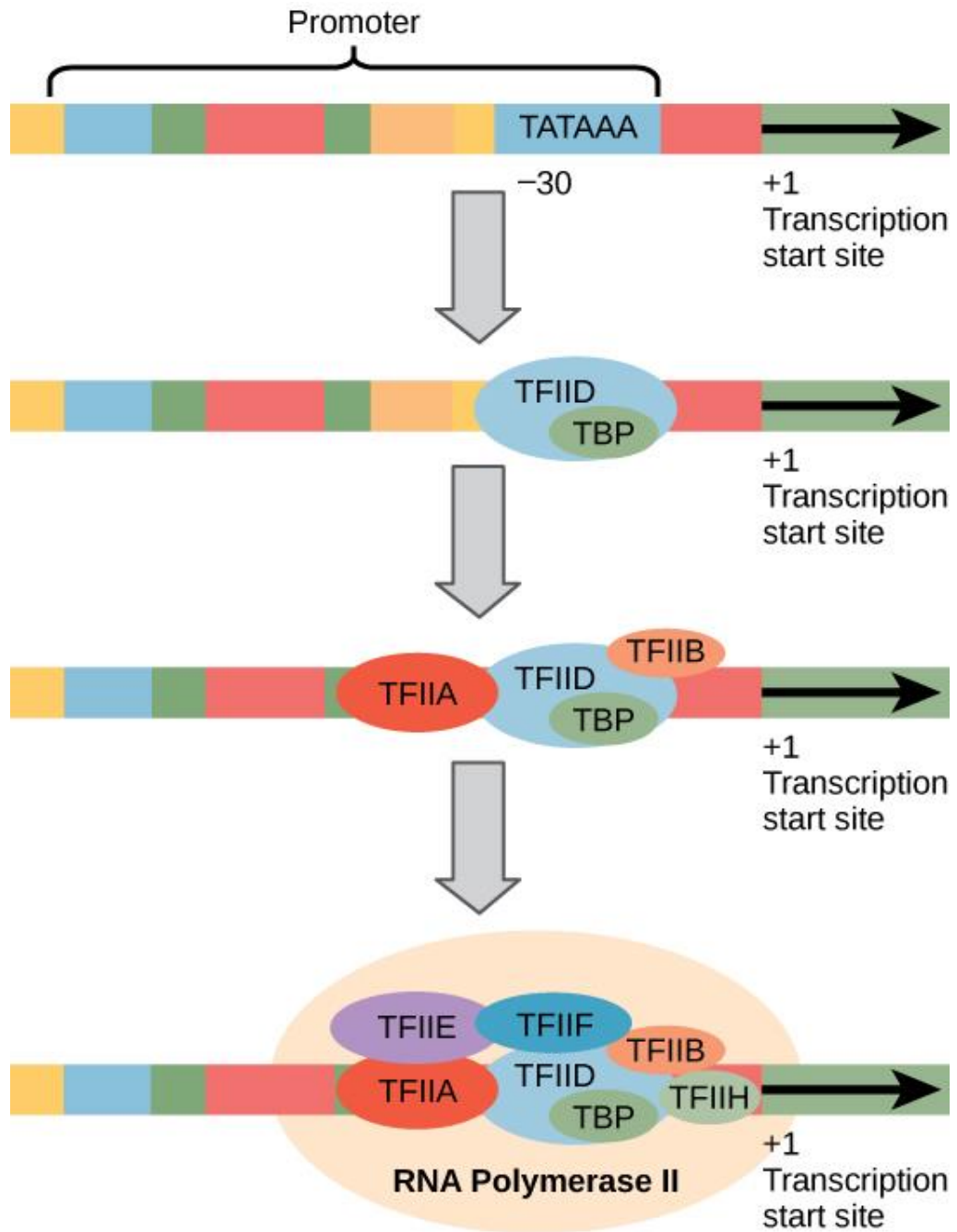
Transcription initiation of a structural gene in eukaryotes.
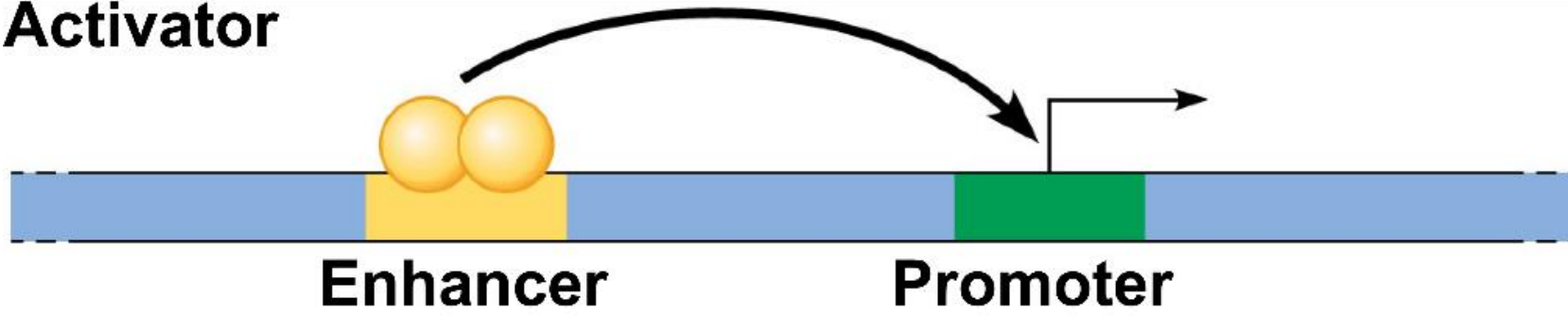
First, the transcription factor TFIID binds to the TATA box, then other transcription factors and RNA polymerase II are attached, and finally, auxiliary factors that trigger transcription. Arrow direction of transcription.
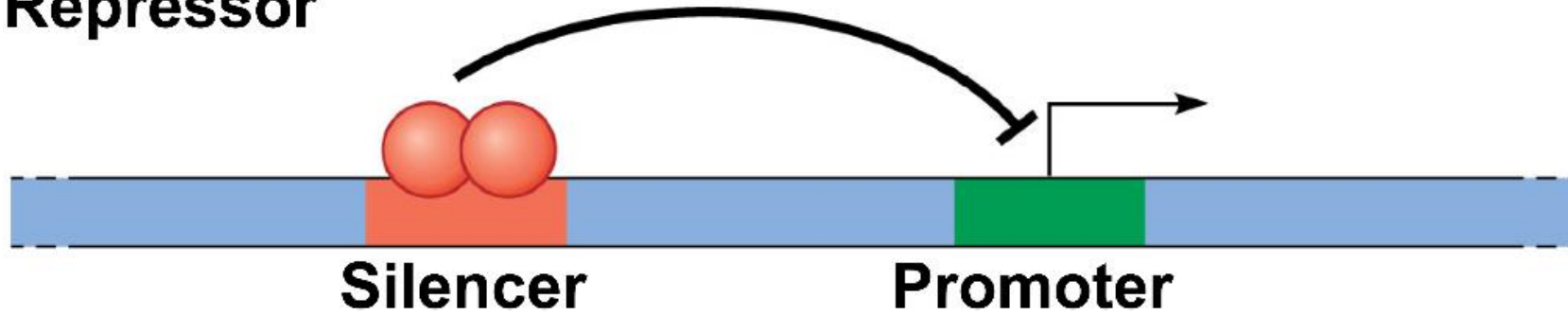
Enhancer

Activator

Mediator

IIF

IIE

IIH

ATP ADP P

TATA Box

P

mRNA

Complex assembly:

TBP → TFIID → IIB → IIA → Pol II

IIF

IIH

IIE

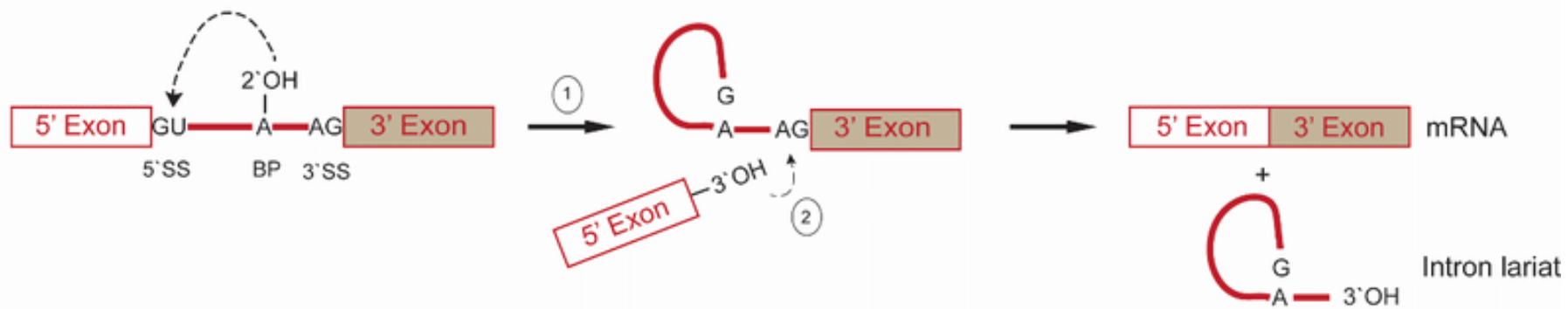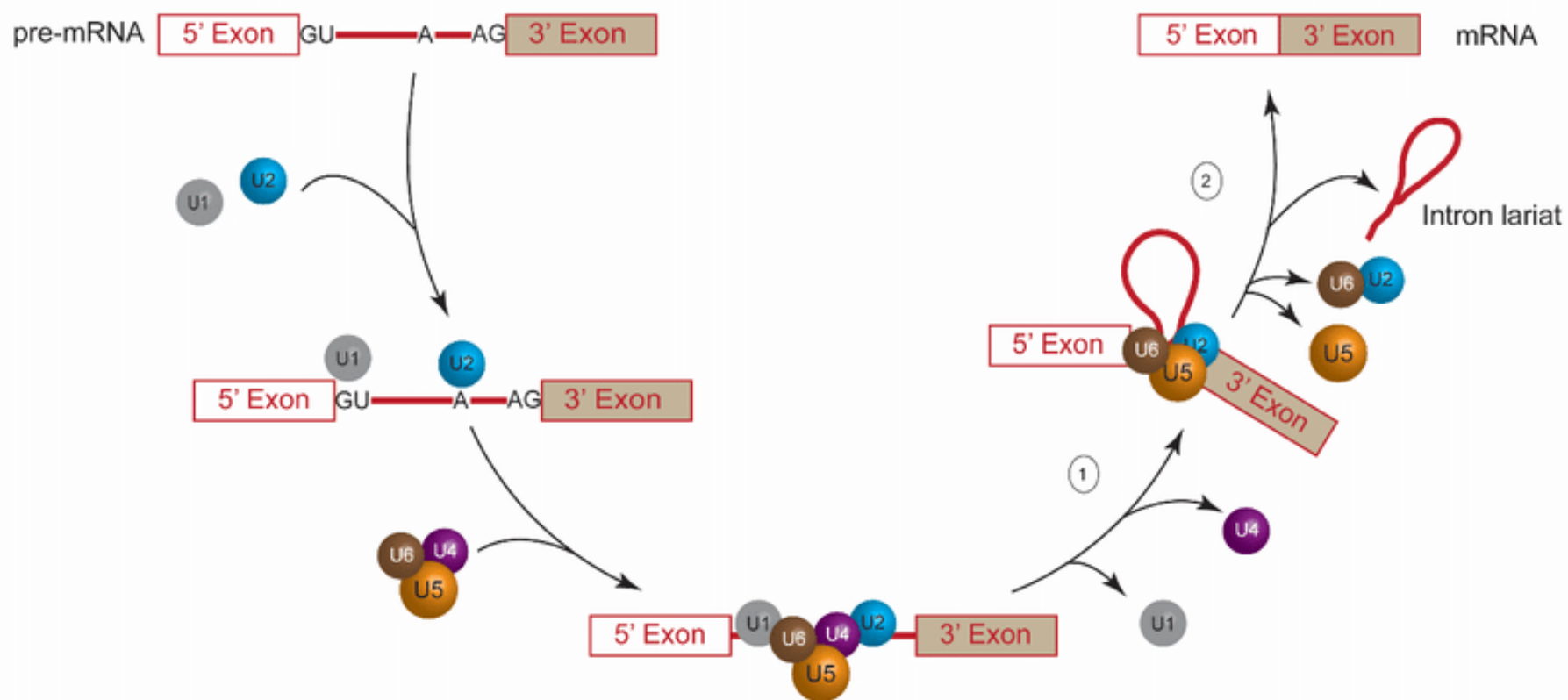Assembled complex

Enhancer and Silencer DNA elements.

- In eukaryotes, most structural genes consist of several discrete coding regions (exons) separated by non-coding regions (introns). Upon completion of transcription of a eukaryotic structural gene, introns are excised from the primary transcription product with the help of enzymes, and exons are sutured together end to end (splicing) to form functional mRNA

- Typically, exons are 150 to 200 nucleotides long and introns are 40 to 10,000 nucleotides long.
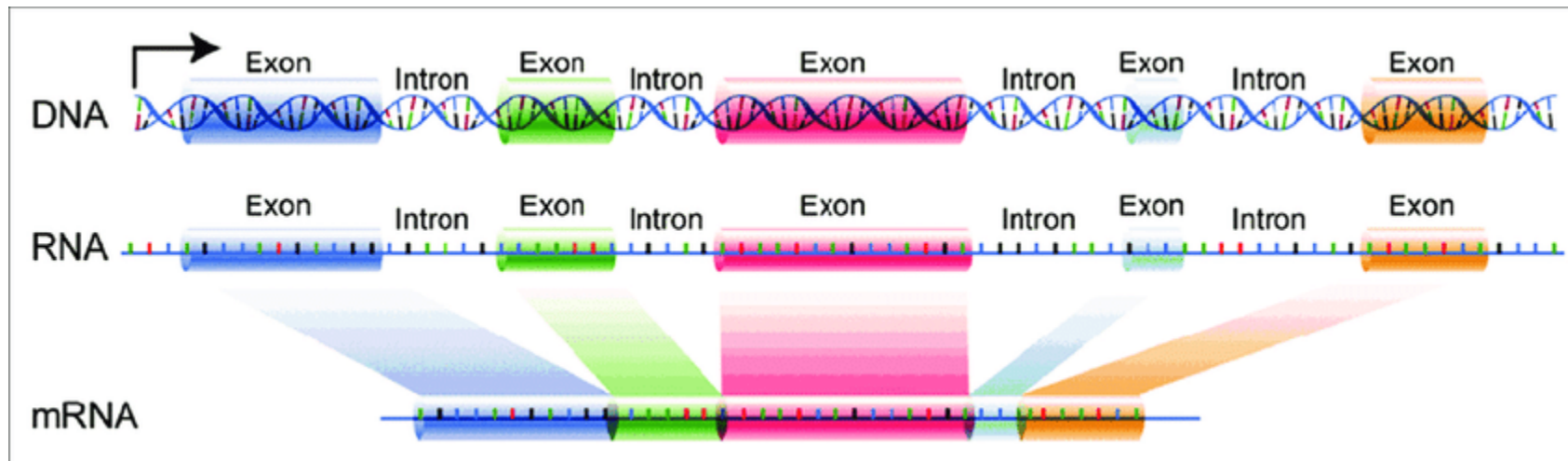

Regulation of mRNA stability.
In eukaryotes, there is also regulation at the level of translation, when ready-made mRNAs are not "allowed" to ribosomes or are destroyed. Other mRNAs can be further stabilized for repeated use.

A

5' Exon GU—A—AG 3' Exon
5'SS   BP 3'SS

mRNA 5' Exon 3' Exon
+
Intron lariat

B

pre-mRNA 5' Exon GU—A—AG 3' Exon

mRNA 5' Exon 3' Exon

Intron lariat

Due to alternative splicing, different products of the same structural gene can be formed in different tissues.
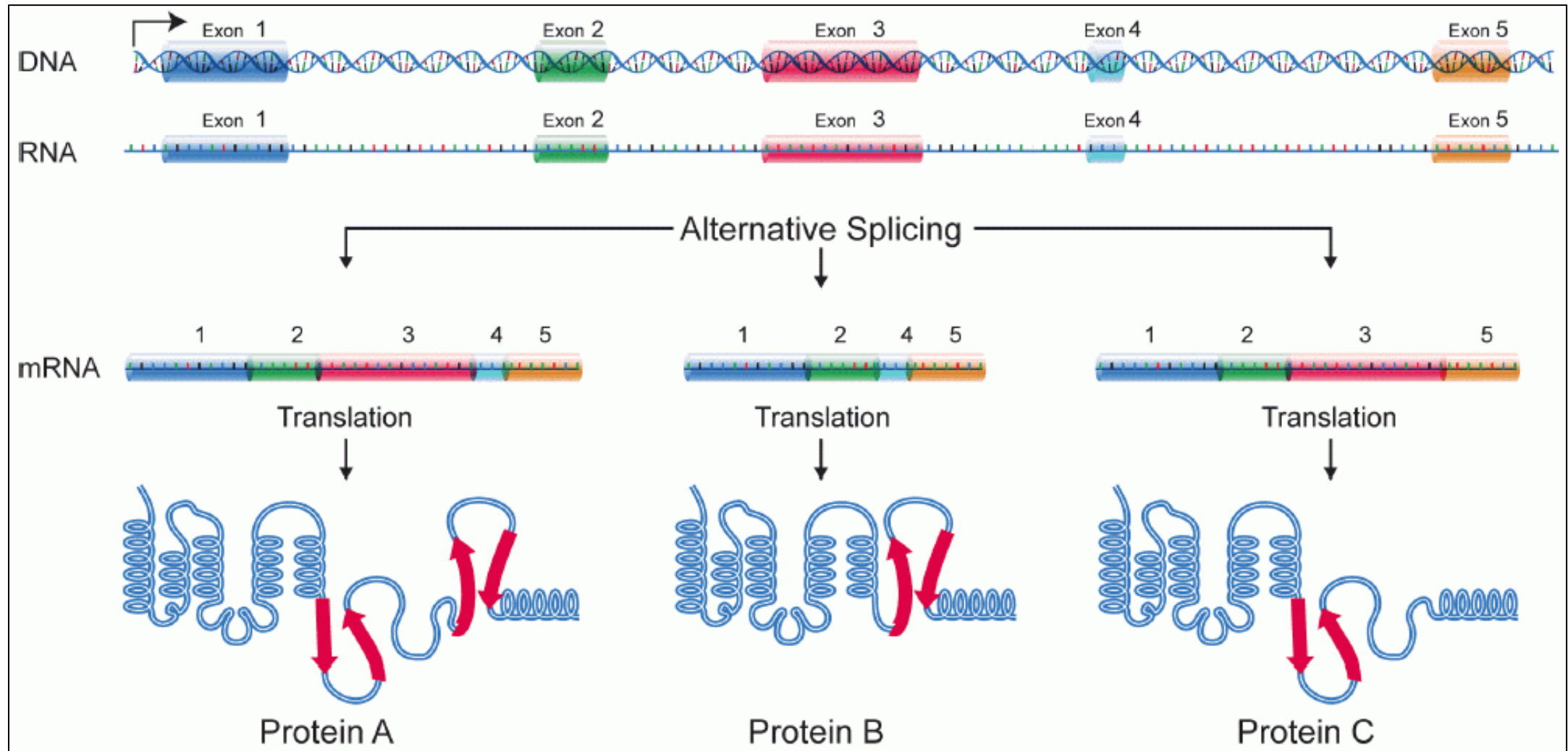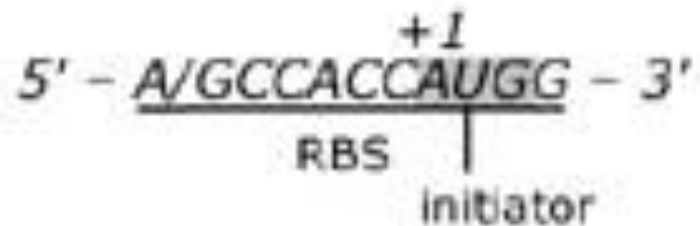
- The Kozak consensus sequence is a nucleotide sequence in eukaryotic mRNA (as well as the corresponding sequence in their genes) that surrounds the start codon and is important for translation initiation.

- Described by American biochemist and molecular biologist Marilyn Kozak in 1986.

Prokaryotic (Shine-Dalgarno sequence)

$+1$

5' – AGGAGGACAGCUAUG – 3'

RBS   spacer   initiator

Eukaryotic (Kozak sequence)

$+1$

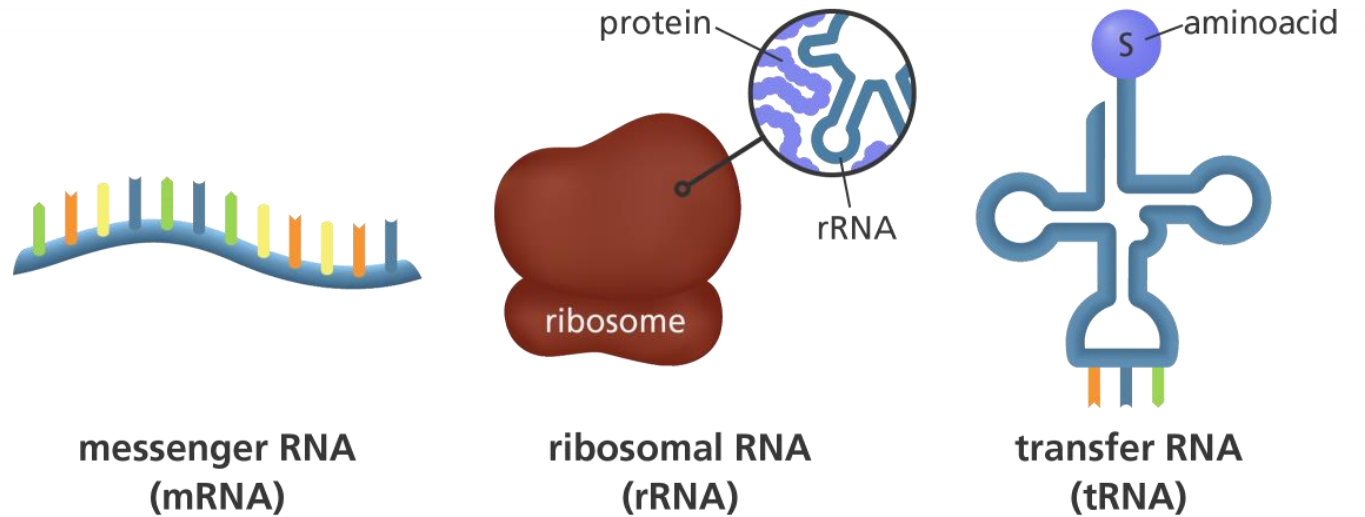5' – A/GCCACCAUGG – 3'

RBS

initiator

Main types of RNA: mRNA, rRNA and tRNA

All of them play an important role in the process of deciphering genetic information.

The synthesis of RNA on a DNA template is called transcription.

In most prokaryotes, transcription of all RNA is carried out using the same RNA polymerase.

In eukaryotes, mRNA, rRNA, and tRNA are transcribed by different RNA polymerases.

protein

rRNA

aminoacid

S

ribosome

messenger RNA
(mRNA)

ribosomal RNA
(rRNA)

transfer RNA
(tRNA)

In an actively functioning cell, approximately 3-5% of total RNA is mRNA, 90% is rRNA, and 4% is tRNA.

- As a result of transcription of discontinuous genes, the primary product is formed - pro-mRNA, which is a complete copy of the gene and contains sections corresponding to both exons and introns.

- The transcription process involves three different types of RNA polymerases that read different genes.

- RNAP-I reads genes encoding the structure of various forms of rRNA (5.8S, 18S, 28S).

- RNAP-II transcribes genes encoding the structure of proteins and some snRNAs.

- Finally, RNAP-III reads the 5S rRNA, transfer RNA, and snRNA genes.
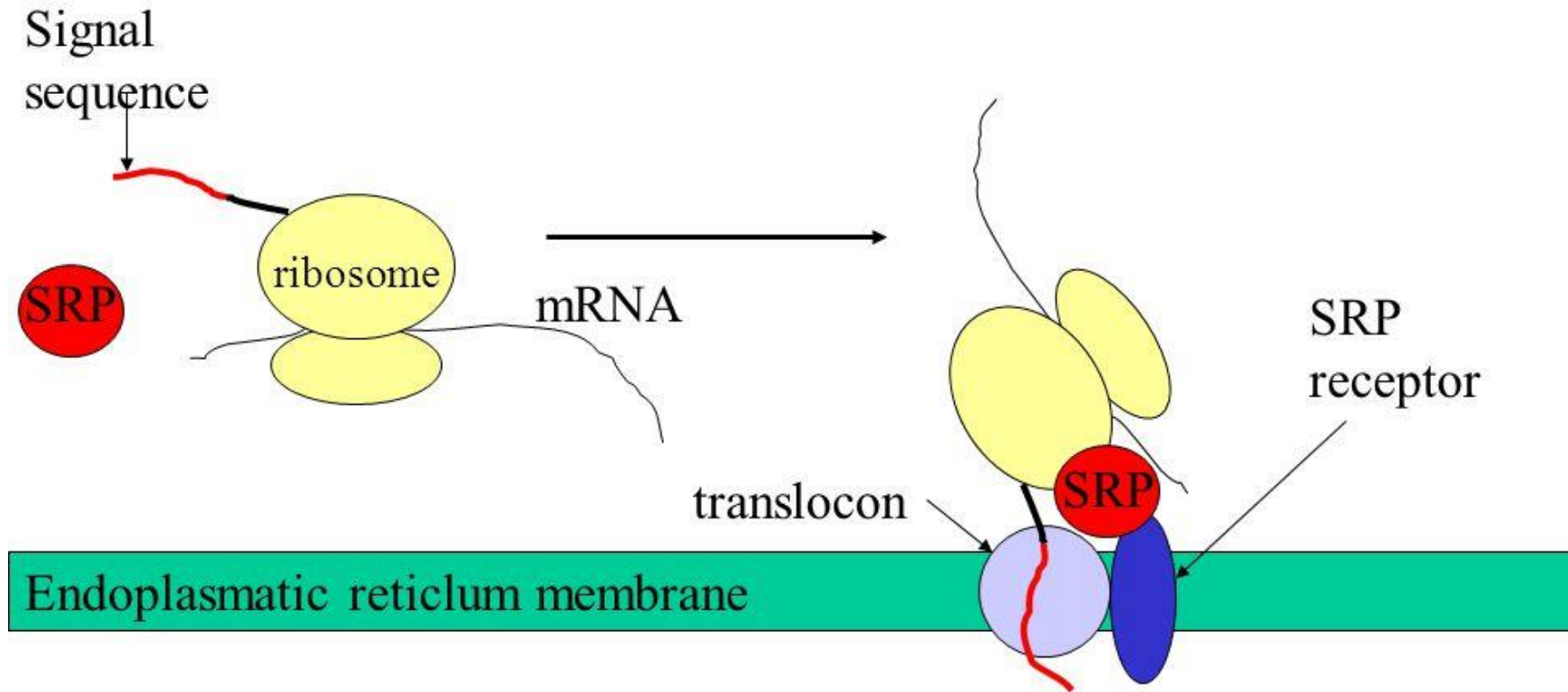
Translation – Initiation
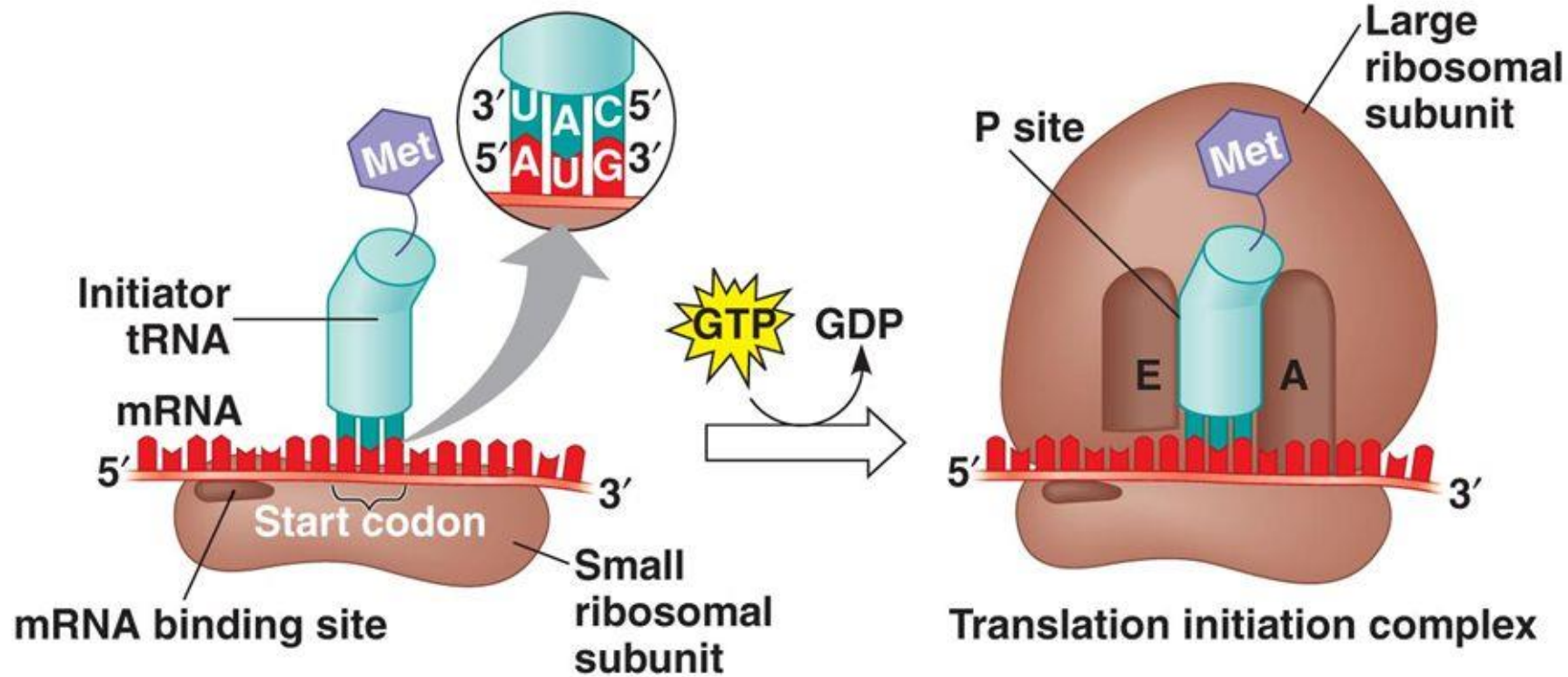
Steps of initiation:

- Small ribosomal subunit binds to mRNA
- An initiator tRNA with the anticodon UAC base pairs with the start codon AUG
- The large ribosomal subunit (rRNA) completes the initiation complex, the energy molecule GTP provides energy for assembly.

4.5S RNA and 7S RNA – a part of signal recognition particle (SRP)

Signal sequence

ribosome

SRP

mRNA

SRP receptor

translocon

SRP

Endoplasmatic reticlum membrane

- SRP recognizes signalling amino acid sequence in the N-terminus of growing polypeptide chain
- Upon signal recognition, ribosome is attached to endoplasmatic reticlum so that the protein, made by ribosome, enters the secretory pathway