

# **DNA transcription – RNA synthesis**

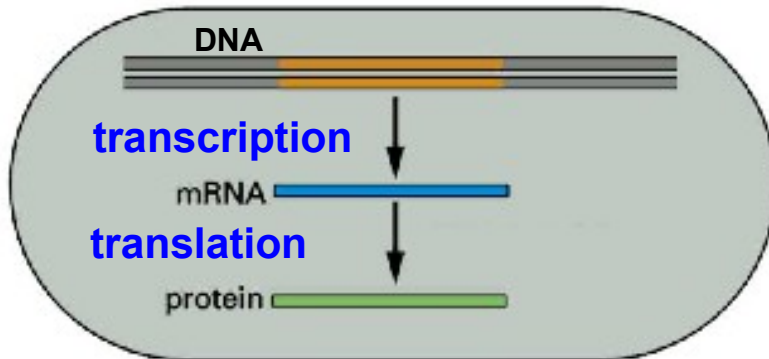
## **Regulation of gene expression**

Biochemistry I  
Lecture 13

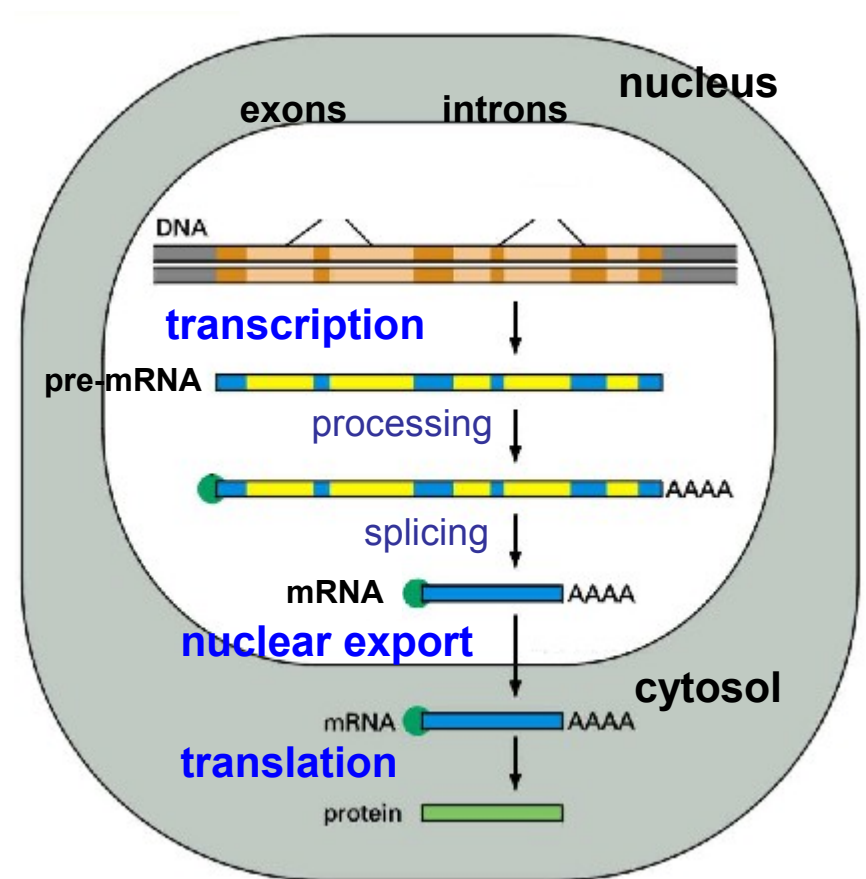
2008 (J.S.)

# Eukaryotic transcription and translation are separated in space and time

## Prokaryotes



## Eukaryotes





# RNA synthesis

Ribonucleoside triphosphates are the substrates for the synthesis.

RNA polymerases (**DNA-dependent ribonucleotidyltransferases**) recognize the nucleotide sequences in the template strands, initiate the synthesis of new chains of RNA **without a primer**, and catalyze the formation of 3'-5' phosphodiester bonds in the complementary transcripts.

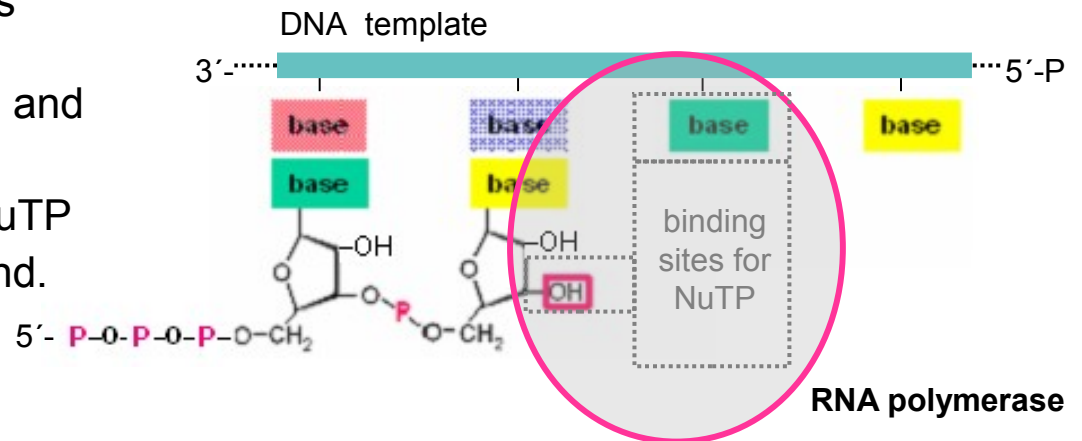
**The nascent RNA chains grow only in the 5' → 3' direction, antiparallel to the direction of the template strand.**

In contradistinction to DNA polymerases, RNA polymerases don't exhibit any nuclease (proof-reading) activity so that they cannot correct mismatches.

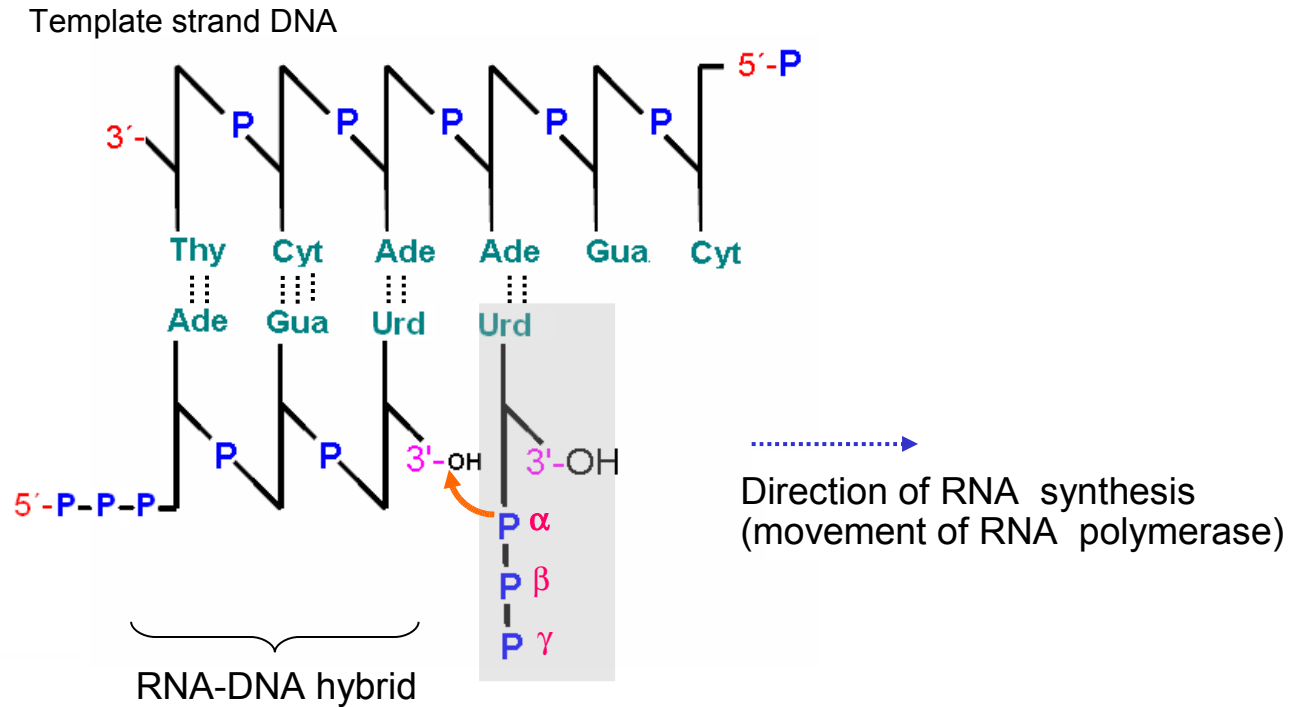
RNA polymerases have binding sites

- for the free 3'-OH group,
- for bases of the template strand, and
- for nucleoside triphosphates.

They cleave  $\beta$ -phosphate bond of NuTP and form 3'-5' phosphodiester bond.



**New 3'-5' phosphodiester bond** originates in the reaction between 3'-OH group of existing chain and  $\alpha$ -5'-phosphate of the incoming nucleoside triphosphate, **diphosphate is released** (complexed with  $Mg^{2+}$  ions).



# RNA polymerases

(DNA-dependent nucleotidyltransferases, transcriptases)

In **prokaryotes**, RNA is synthesized by a single kind of RNA polymerase.

RNA polymerase from *Escherichia coli* consists of five subunits of four kinds, one of which is the  $\sigma$  factor that helps find a promoter site where the transcription begins (and then dissociates from the rest of the enzyme).

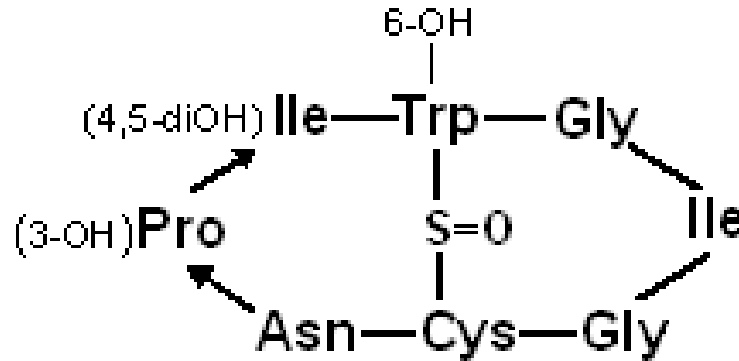
In **eukaryotes**, the nucleus contains **three types of RNA polymerase**.

The mechanism of their action is the same, but they differ in binding onto different promoters (template specificity), location in the nucleus, and also in susceptibility to inhibitor  $\alpha$ -amanitin. RNA polymerases contain from 8 to 14 subunits ( $M_r > 500\,000$ ).

In the mitochondrial matrix, there is the fourth type – **mitochondrial RNA polymerase**.

RNA polymerase	Nuclear location	Primary transcripts
<b>pol I</b>	<u>nucleolus</u>	<b>pre-rRNA</b> 45 S
<b>pol II</b>	nucleoplasm	<b>pre-mRNAs</b> , some snRNAs
<b>pol III</b>	nucleoplasm	<b>pre-tRNAs</b> , rRNA 5 S, some snRNAs

*Amanita phalloides* (the death cup) produces  **$\alpha$ -amanitin** that blocks the elongation phase of RNA synthesis



**$\alpha$ -Amanitin** is a cyclic octapeptide, in which the sulfinyl group (oxidized sulfanyl group of the cysteinyl residue) is attached to the indole ring of the tryptophyl residue.

It is an effective inhibitor of eukaryotic RNA polymerases II and III, namely that of the polymerase II.

# Transcription of DNA

is a three-phasic process consisting of initiation, elongation, and termination.

Transcription starts at **promoters** on the DNA template.

Promoters are sequences of DNA that direct the RNA polymerase to the proper initiation site for transcription. Each of the three types of RNA polymerase has distinct promoters.

Promoters are mostly in the normal upstream position to the initiation site.

The effectiveness of promoters can be regulated (increased or restrained) by specific DNA sequences called **enhancers** or **silencers** that may be distant up to 2000 base pairs from the promoter either upstream or downstream.

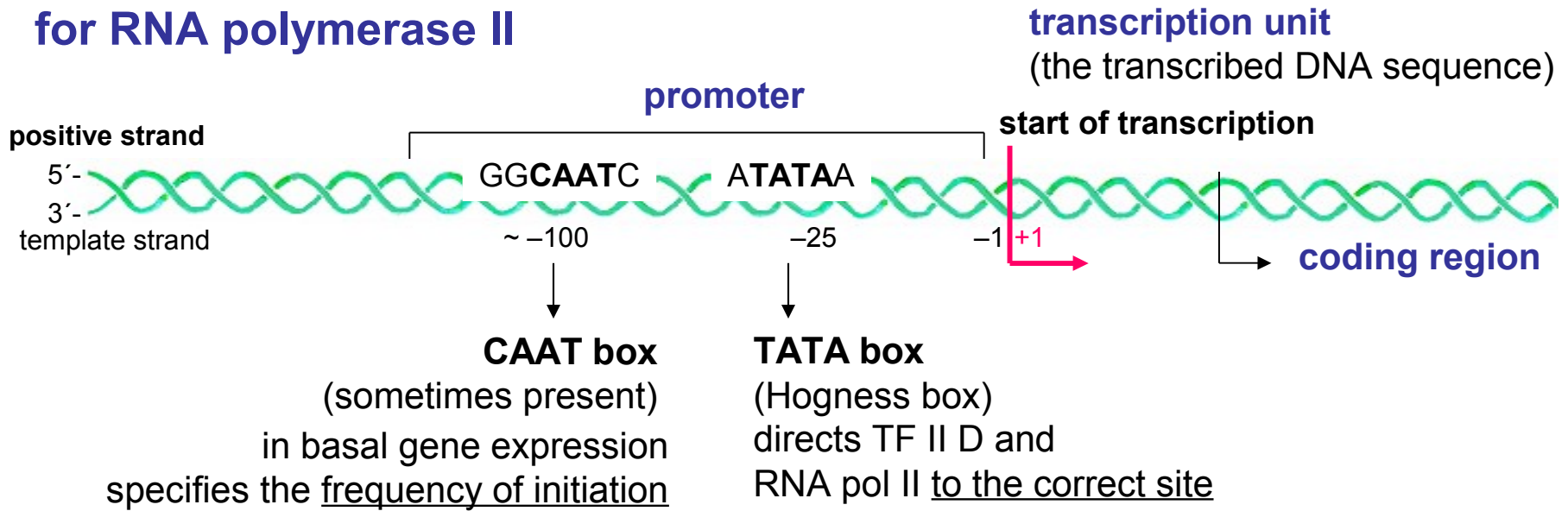
Promoters and enhancers are referred to as **cis-acting elements**, because they are sequences of the same molecule of DNA as the gene they regulate.

The DNA sequences of cis-acting elements are binding sites for proteins called **transcription factors**.

If those factors are encoded by a gene on a DNA molecule other than that containing the gene being regulated, they are called **trans-acting factors**.



# Eukaryotic promoter site for RNA polymerase II

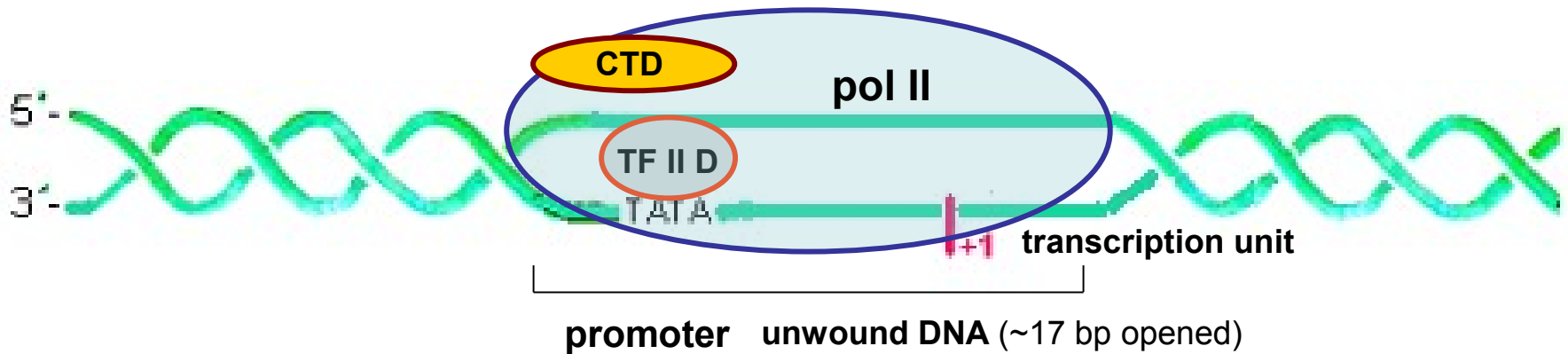


Polymerase II and transcription factors bound onto the promoter form a complex called the **basal transcription apparatus**. It regulates basal gene expression. Genes that are regulated wholly in this way are **constitutively expressed genes** (products of which are most constitutive proteins).

Specifically regulated expression of numerous genes is mediated by various **gene specific transcription factors**. Those proteins (coactivators, corepressors, transcoactivators, etc.) bind to regulatory DNA sequences distant from promoters. The basal transcription apparatus is thus regulated through direct or mediated contact with the gene specific transcription factors. See "Regulation of gene expression".

# Transcription initiation

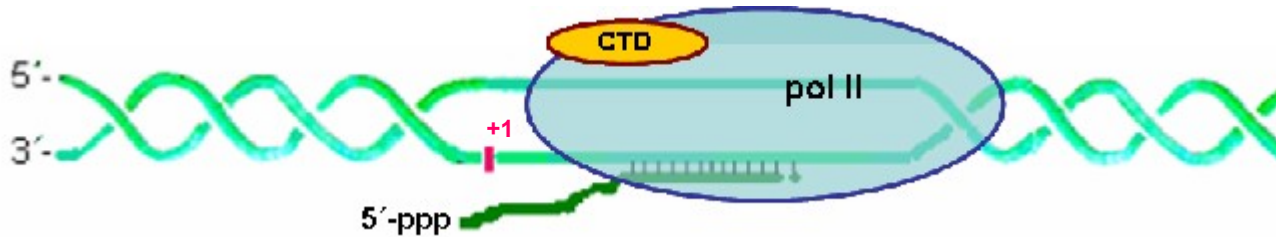
Initiation begins with the binding of **TF II D** (transcription factor D for pol II) to the TATA box. TF IID provides docking sites for binding of other transcription factors. One of those factors is an ATP-dependent **helicase** that separates the DNA duplex for the **polymerase II**, the last but one component of the basal transcription apparatus. Pol II contains an **unphosphorylated carboxy-terminal domain (CTD)**.



Polymerase II with its unphosphorylated CTD then slides to the start of transcription and initiate transcription by producing short transcripts consisting of not more than 20 – 25 nucleotides.

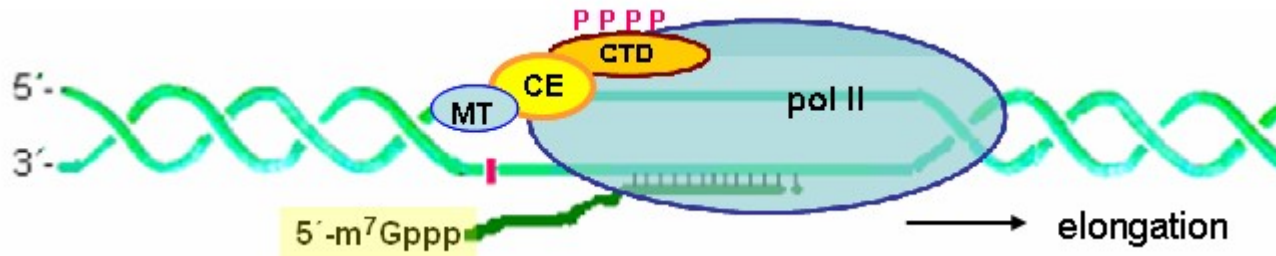
After the transcription is initiated, most transcription factors are released from pol II.

## Switch from initiation to elongation is driven by phosphorylation of carboxy-terminal domain of pol II

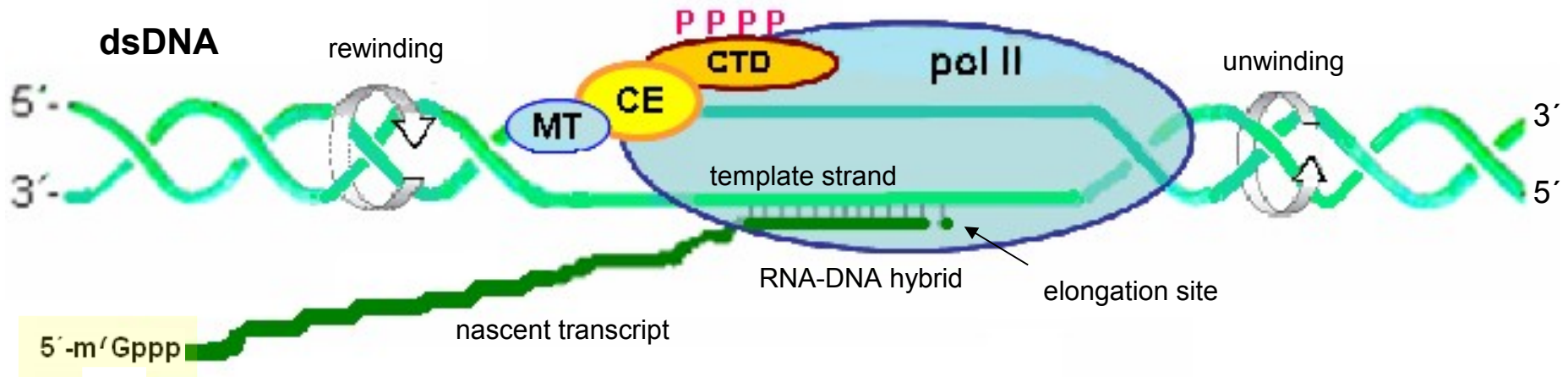


**CTD is then phosphorylated.** The resulting change in conformation of pol II (from pol II A to pol II O form) enables binding of capping enzyme (**CE**) to pol II and methyltransferase (**MT**) to CE. Both those enzymes modify the 5'-end of the nascent transcript to **5'-m<sup>7</sup>Gppp-cap** required for the further progress in transcription.

The phosphorylated CTD of pol II has a central role in cotranscriptional RNA processing, because it also binds, in addition, splicing factors and factors responsible for the final polyadenylation of the transcripts.

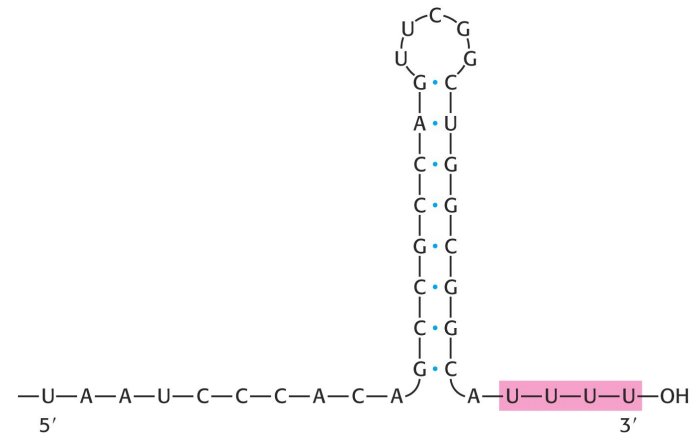


# Elongation phase



# Termination

In **prokaryotes**, termination signals usually contain a palindromic GC-rich region and an AT-rich region. Thus the mRNA transcripts of this DNA palindrome can pair to form a hairpin structure with a stem and loop followed by a sequence of more uracil base – RNA transcripts end within or just after them

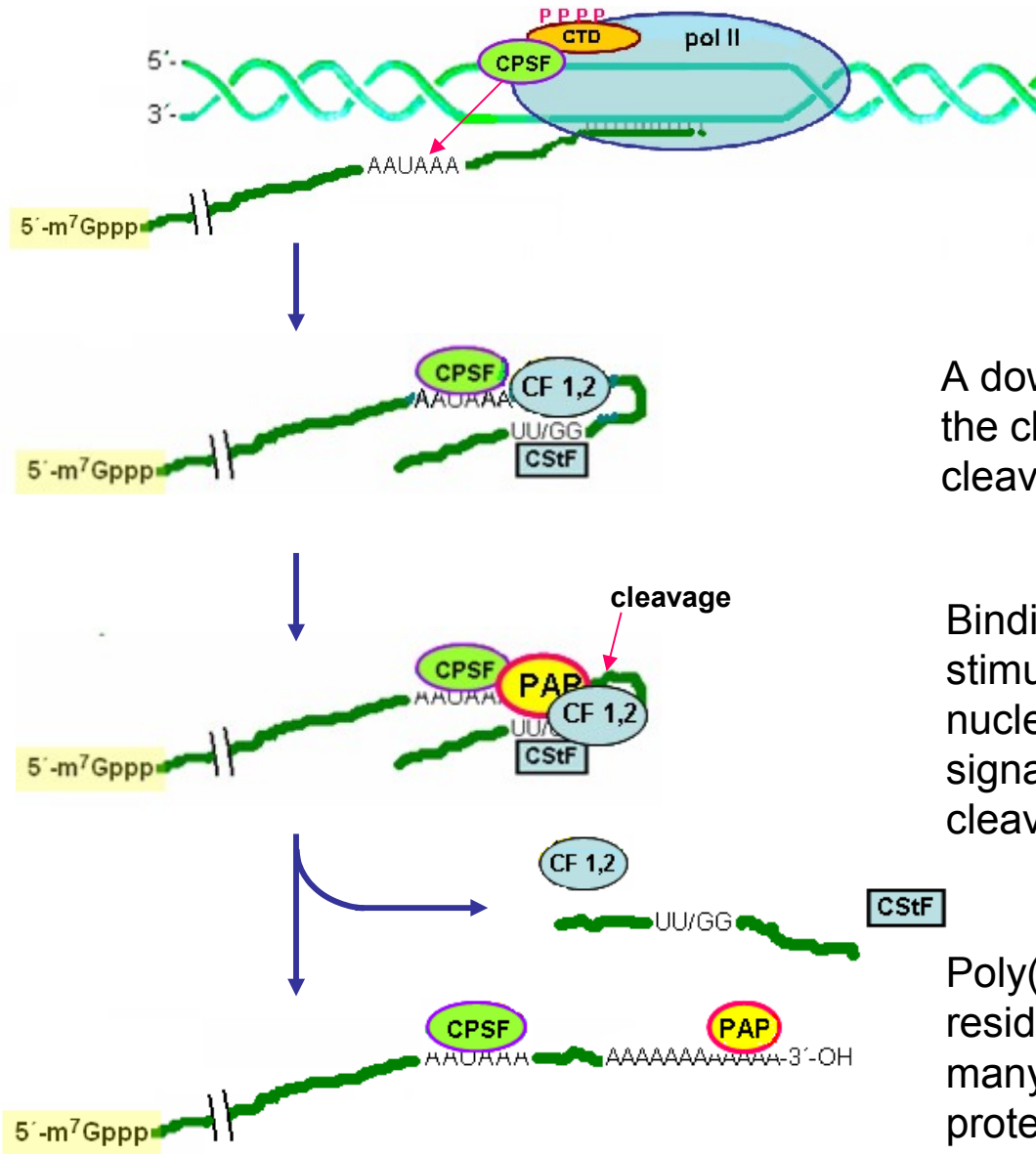


## In eukaryotes,

no perspicuous termination signal has been found.

**Transcripts produced by DNA polymerase II** are released from the transcription apparatus after the polyadenylation signal AAUAAA and the GU- or U-rich sequence that is able to bind cleavage stimulation factor (CStF) had been transcribed. The terminal sequences of the transcripts are decomposed in the course of 3'-polyadenylation (not encoded by template DNA).

# Polyadenylation of transcripts



**Cleavage-and-polyadenylation specificity factor (CPSF) binds onto an polyadenylation signal AAUAAA.**

It is not quite clear when transcripts are released from the transcription apparatus.

A downstream GU- or U-rich sequence binds the cleavage stimulation factor (CStF) and cleavage factors (CF 1,2), a loop is formed.

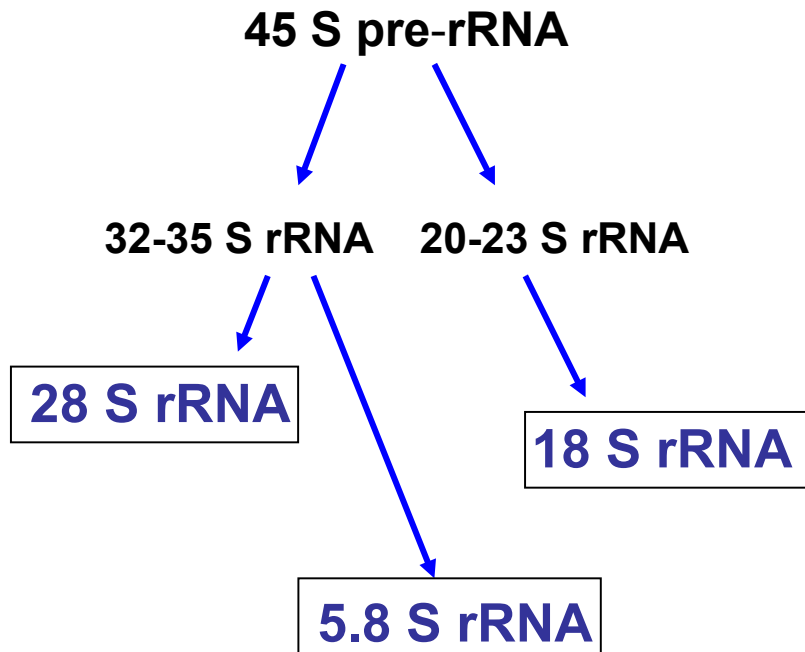
Binding of **poly(A) polymerase (PAP)** then stimulates cleavage at a site about 20 nucleotides downstream the polyadenylation signal. The cleavage factors are released, the cleaved RNA chain degraded.

Poly(A) polymerase adds 12 adenylate residues, elongation is provided by transfer of many short poly(A) chains from poly(A)-binding protein.

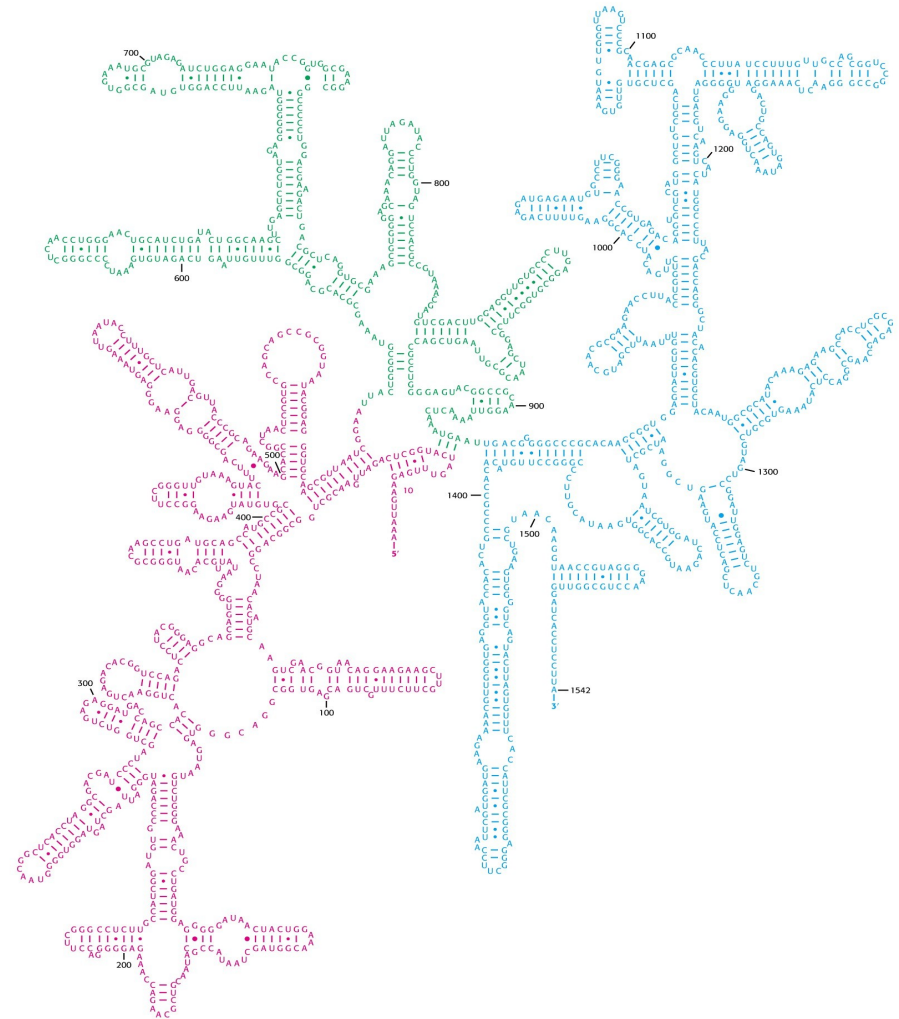
The transcription products of all three eukaryotic polymerases are processed before their export from the nucleus

Precursors of rRNA are cleaved in functional rRNA types.

RNA polymerase I transcribes  
45 S pre-rRNA within the nucleolus:



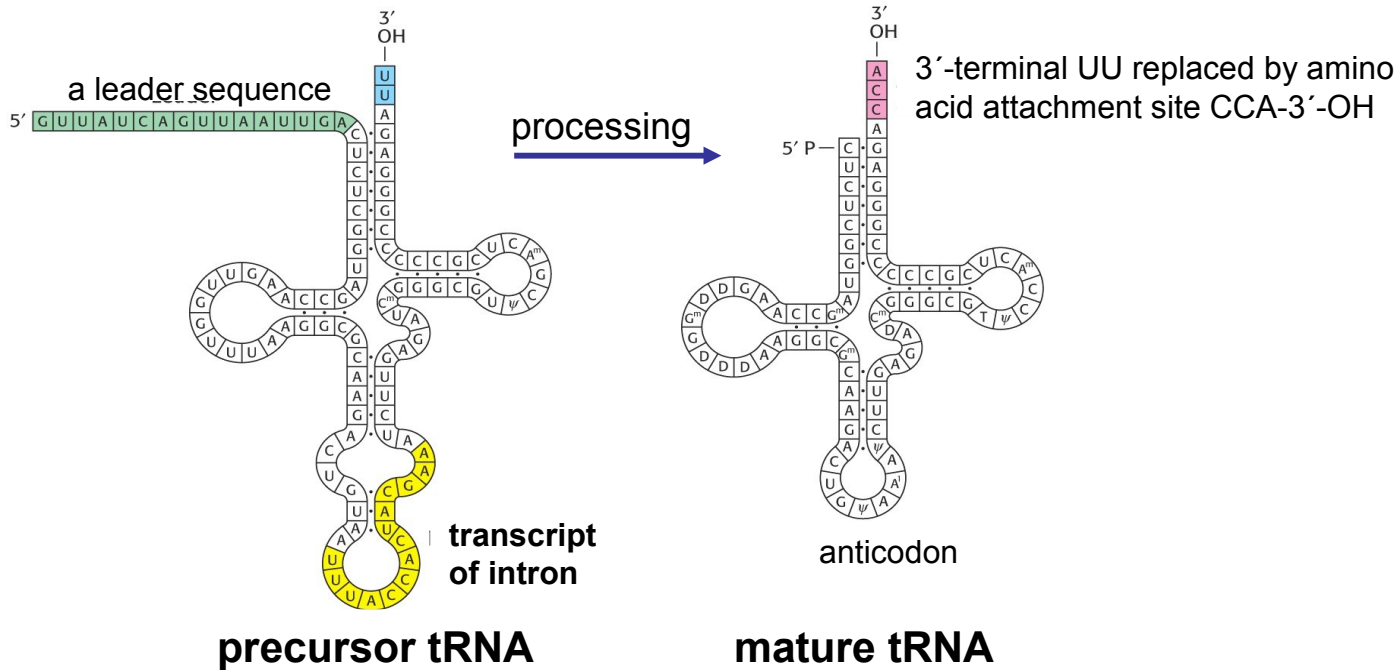
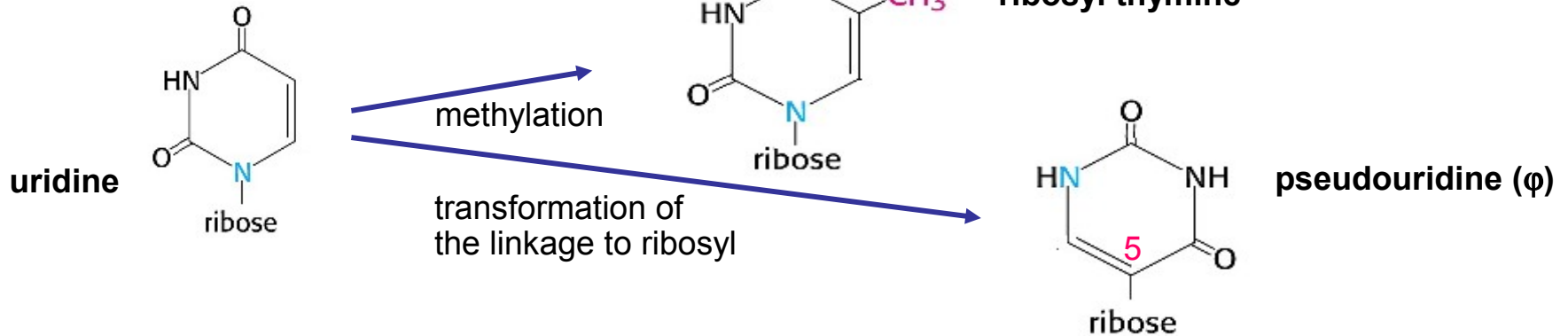
Ribosomal RNA folding pattern (rRNA 18 S)



The fourth ribosomal RNA **5 S rRNA** is transcribed by RNA polymerase III within the nucleoplasm.

# Examples of tRNA processing:

Modification of some bases





## Cotranscriptional and posttranscriptional processing of transcripts (pre-mRNA) produced by RNA polymerase II

Primary transcripts of genes transcribed by RNA pol II (precursor mRNAs) undergo processing, **mostly before their transcription is finished**:

The **7-methylguanosine "cap"** is attached to the 5'-end triphosphate.

The transcripts of non-coding sequences of the gene (introns) are cut off and the transcripts of coding sequences (exons) spliced, the process is called **splicing**.

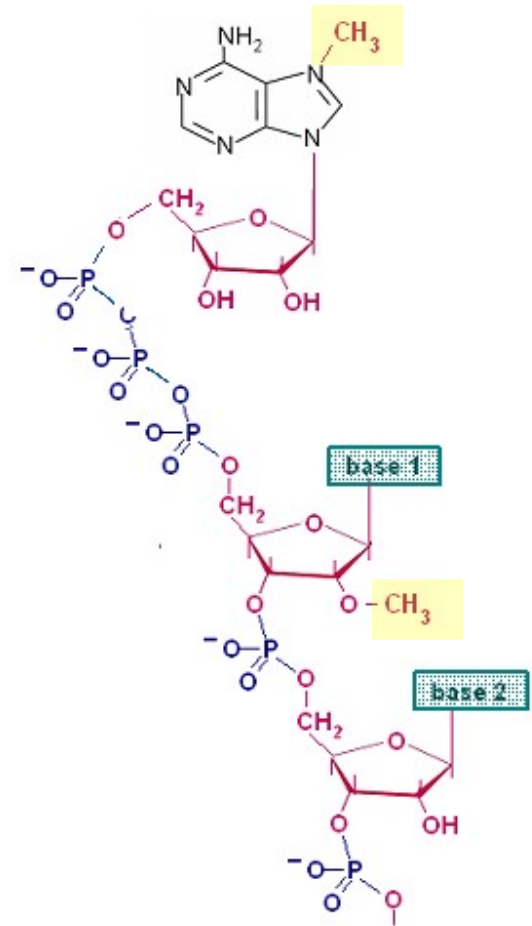
Termination of transcription is connected with the **adding of a polyadenylate chain** to the 3'-end (after cleavage of the terminal sequence of the primary transcript).



In some mRNAs, the base sequence is altered after transcription by processes other than RNA splicing. Those processes are called **RNA editing** and are not very rare. E.g. cytidine residue may be deaminated to uridine, adenosine to inosine.

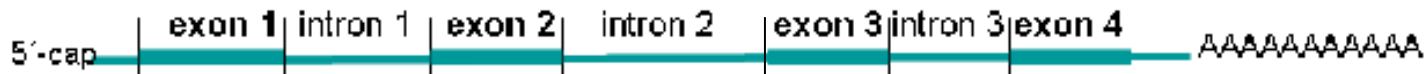
Processed mRNAs bind certain kinds of proteins and form so complexes called **messenger ribonucleoproteins (mRNP)** that are exported through nuclear pore complexes into cytoplasm.

The 7-methylguanosine 5'-cap prevents mRNA against 5'-endonucleases and it is also the marker recognized in proteosynthesis.



### Splicing schematically:

pre-mRNA

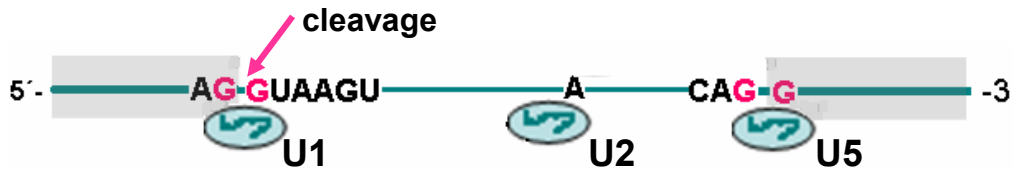
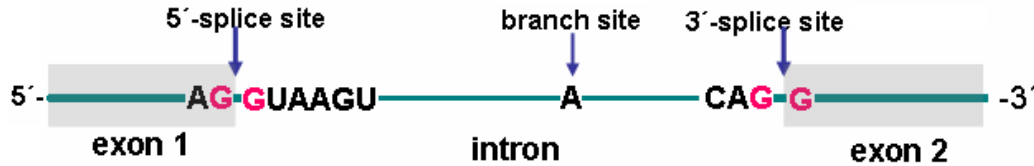


mRNA

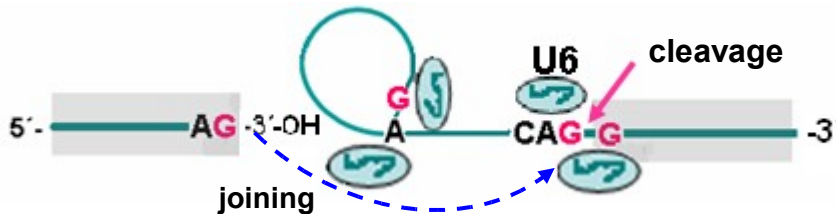
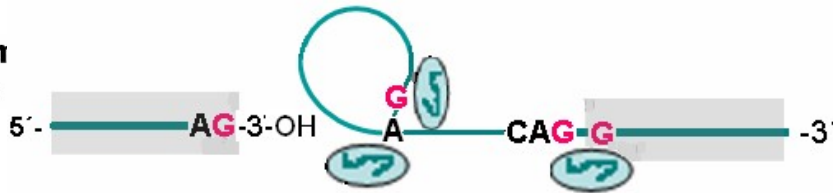


# Splicing

Nucleotide sequences determine the splice sites:

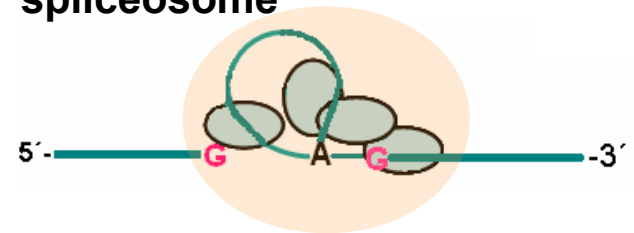


Lariat form of intron

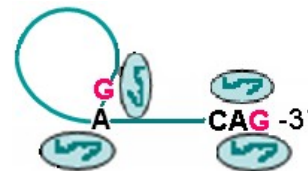


5'-2'-phosphodiester bond between 5'-end of the intron and branch site Ado forms a lariat

spliceosome



Spliced product



Excised intron sequence is degraded in the nucleus

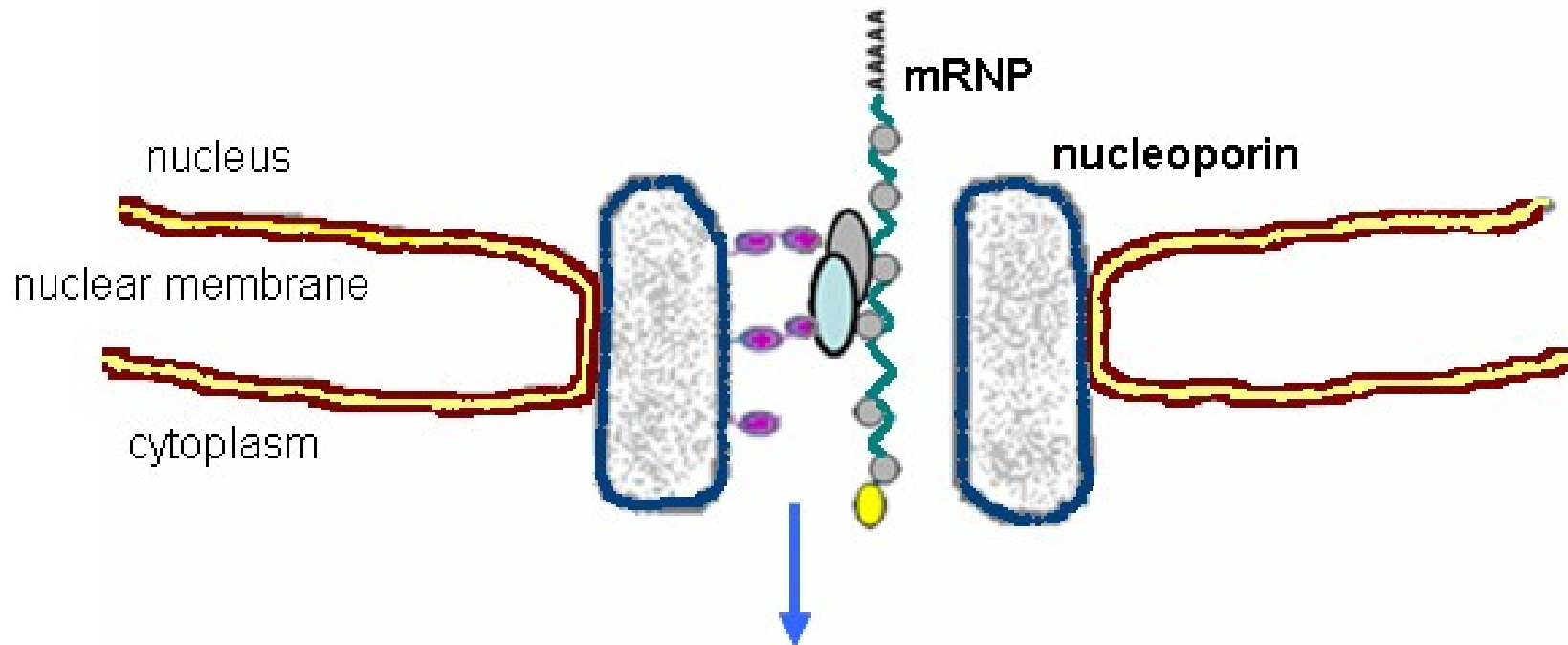
There are many types of small RNA molecules with fewer than 300 nucleotides in the nucleus - **small nuclear RNAs (snRNAs)**. A few of them are essential for splicing pre-mRNA. They associate with specific nuclear proteins to form complexes called **small nuclear ribonucleoprotein particles (snRNP, "snurps")**.

**Small nuclear ribonucleoprotein particles (snRNPs)  
in the splicing of mRNA precursors**

snRNP	Size of snRNA (number of nucleotides)	Function
U1	165	Binds the 5' splice site and then the 3'-splice site
U2	185	Binds the branch site and forms part of the catalytic center
U5	116	Binds the 5' splice site
U4	145	Masks the catalytic activity of U6
U6	106	Catalyzes splicing

During the splicing of pre-mRNA, the processed mRNA, snRNPs U1, 2, 4, 5, 6, and other protein splicing factors form large assemblies (about 60 S) termed **spliceosomes**.

## Export of messenger ribonucleoprotein particles (mRNPs) through the nuclear pore complex



For the transport, **messenger ribonucleoprotein particle (mRNP)** associates with the heterodimer known as the **general mRNA export receptor (expartin 1)**.

The **nuclear pore complex** consists of proteins **nucleoporins**, which contain Phe-Gly repeats and zinc-finger domains. These structures provide transient docking sites for the complex export receptor - mRNP traversing the nuclear pore "basket".

# Regulation of gene expression

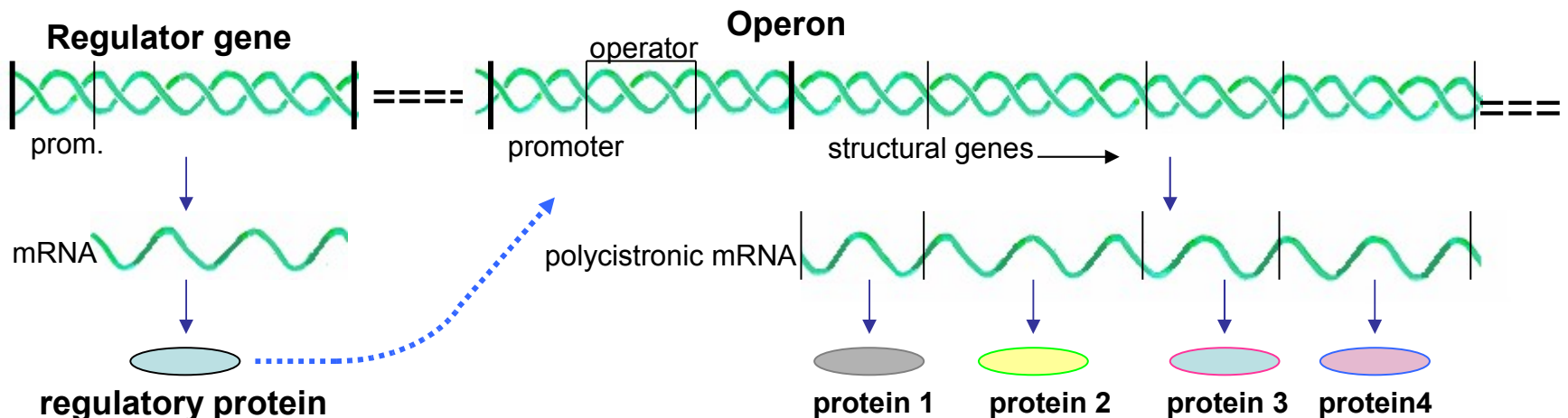
**Gene expression** is the term that involves **conversion of the genetic information** encoded by a gene **into the final gene product**, i.e. a protein or a functional RNA (rRNA, tRNA).

Control of gene expression in prokaryotes differs from that in eukaryotes distinctly.

## Gene expression in prokaryotes

In prokaryotes, gene activity is controlled foremost at the level of transcription, at its initiation.

The structural genes are usually grouped together in **operons**, which are transcribed from one promoter controlled by a regulatory protein.



## Negative control of transcription in prokaryotes

is based on the existence of regulatory proteins named **repressors** (products of specific regulatory genes) that bind onto specific operator sequences within the promoters and **prevent from binding RNA polymerase** and initiation of transcription of genes of the regulated operon.

### Repression of gene expression

In repressible transcriptions, repressor proteins are produced usually in their inactive forms. They are able to bind onto specific operons and act as **active repressors only in the presence of co-repressors**, allosteric activators, which change their conformation.

Repressible operons are oft those, which provide enzymes for biosynthetic pathways that can be blocked by the presence of product of the synthesis.

### Induction of gene expression

In inducible transcription, the regulatory protein repressor is produced constitutively and is bound to the operator, transcription of the gene cannot occur: **inducers are allosteric effectors, which bind the repressors and lower their affinity for the operator**. In the presence of the inducer, the repressor leaves the operator and transcription of the gene begins. So induction in prokaryotes is **derepression** in fact.

Inducible operons are mostly those providing enzymes of catabolic pathways (operon expressed only in response to the presence of substrate).



## Examples of negative control in prokaryotes:

### Induction of *E. coli lac* operon by lactose

The enzymes of glucose metabolism are constitutive (synthesized all the time). If lactose appears in the medium lacking in glucose, cells begin to produce enzymes which are able to metabolize lactose ( $\beta$ -galactosidase, permease, and transacetylase). Inductor is **allolactose** (formed by spontaneous isomerization of lactose) that **binds to the *lac* operon repressor and inactivates it.**

### Repression of *E. coli trp* operon by tryptophan

The genes encoding five enzymes essential for the biosynthesis of tryptophan are included in *trp* operon. Repressor of *trp* operon is expressed constitutively, but in its inactive form that is activated by binding tryptophan. Thus the biosynthesis of tryptophan is inhibited, if there is sufficient tryptophan concentration within the cell. **Tryptophan acts as corepressor of *trp* operon that prevents from its transcription.**

## Positive control of transcription in prokaryotes

Regulatory proteins can act sometimes as **activators**, which bind near the promoter and **support binding of RNA polymerase** to the promoter.

### An example of positive control:

In *E. coli*, transcription of *lac* operon induced by lactose increases approximately 50 times in the rate in the presence of "**catabolite gene activator**" protein (**CAP**) complexed with **cAMP**; however, concentration of cAMP in *E. coli* is high **in the absence of glucose**. The consequence is that lactose can be effectively metabolized only if there is insufficient supply of glucose – the preferred nutrient.

**Glucose** lowers the concentration of cAMP. cAMP then cannot bind the CAP and transcription of operon providing enzymes of lactose catabolism is insufficient even in the presence of the inducer lactose.

This type of positive control of *lac* operon is known as **catabolite repression**, because it disables in the presence of glucose.

# Regulation of gene expression in eukaryotes

Let us remember the differences between gene expression in prokaryotes and eukaryotes. In eukaryotes:

- Nuclear DNAs are highly condensed in chromosomes and, in addition, they interact tightly with histones forming nucleosomes.
- Each gene has its own promoter, there are no operons.
- In primary transcripts, the transcripts of introns are included that have to be excised.
- Transcription and translation are separated both in space and in time.

The control of eukaryotic gene expression occurs principally at the level of transcription. However, there are numerous other ways of control:

## Regulation at the level of

- 1 chromatin and DNA,**
- 2 transcription,**
- 3 processing of primary transcripts,**
- 4 translation and posttranslational processing.**

# 1 Regulation at the level of chromatin and DNA

## Control of the gene accessibility for transcription

Chromatin of chromosomes occurs in two kinds, either as condensed heterochromatin (the included genes are not active transcriptionally) or diffused euchromatin, the genes of which are transcribed. Each cell of the organism (with a few exceptions) contains the same complement of genes. However, the **changes in chromatin structure type** occurring **in development and differentiation** of tissues and cell types result in differential gene expression.

## Chromatin remodeling

are those changes in organisation of dsDNA in chromatin fibres that are required for initiation of transcription. Various mechanisms of remodeling exist.

E.g., unwinding of dsDNA segments from nucleosomes depends on both hydrolysis of ATP and **covalent modification of histones** (acetylation of  $\epsilon$ -amino groups of lysyl residues at N-ends of histones H2A, H2B, H3, and H4). on se

## Methylation of DNA

Methylation of cytosine base of DNA to 5-methylcytosine occurs oft in the GC-rich sequences near gene promoters. This modification of DNA is catalyzed by specific methylases, **genes containing 5-methylcytosines are transcribed less easily** than those non-methylated.

**Example:** The genes for  $\alpha$ - and  $\beta$ -globin chains are methylated in non-erythroid cells that cannot synthesize haemoglobin. In erythroblasts and reticulocytes (precursors of red blood cells), those genes are not methylated.

## Selective gene rearrangements

The coding segments of DNA can recombine within the particular gene or may associate with other genes within the genome.

**Example:** Recombination of gene segments type V, J, and C of genes for immunoglobulins is the cause of **vast diversity of specific antibodies**.

## Amplification of genes

Eukaryotic genes can be amplified during development or in response to drugs.

Certain parts of chromosomes is repeatedly replicated during particular cell cycle.

Newly synthesized DNA is excised in the form of small, unstable chromosomes (called double minutes) that are incorporated into other chromosomes.

Amplification occurs normally due to mistakes in DNA replication or cell division.

However, under appropriate conditions, these extra rounds of replication can become "frozen" in the genome.

**Example:** In patients receiving methotrexate (inhibitor of dihydrofolate reductase) can develop **drug resistance** by increasing the number of genes for dihydrofolate reductase by gene amplification.

## 2 Regulation at the level of transcription

The most important and fundamental element in the initiation of transcription is the **promoter**, on which the **basal transcription machinery complex** (basal transcription factors and RNA polymerase) is assembled.

### Basal control of transcription

seems to be common to all genes. It includes binding of **basal transcription factors** to the promoter or closely adjacent sites. Some of those factors determine by binding to GC and CAAT boxes how frequently transcription is to occur.

### Specific control of gene expression

Gene-specific or tissue-specific expression depends on

- **regulatory DNA sequences** within the same DNA molecule (*cis*- elements), which can influence transcription even when separated by thousands of base pair from promoter - enhancers, silencers, hormone response elements, and
- **specific transcription factors** – proteins originating from genes presumably located on different chromosomes (*trans*-acting elements), which don't bind to the promoter or closely adjacent DNA sites.

## Regulatory DNA sequences

can either increase (**enhancers**) or decrease (**silencers**) the rate of transcription of eukaryotic genes. This effect is **mediated by specific transcription factors**. Enhancers bind transactivators or coactivators, silencers bind corepressors.

**Hormone response elements (HRE)** are regulatory DNA sequences that bind complexes of hydrophobic hormones (steroid and thyroid hormones, retinoates) with their intracellular receptors. They act as enhancers or silencers.

## Specific transcription factors

are proteins, which bind to regulatory DNA sequences remote from the promoter. and act as activators (**transactivators** or **coactivators**) or repressors (**corepressors**) of transcription of specific gene.

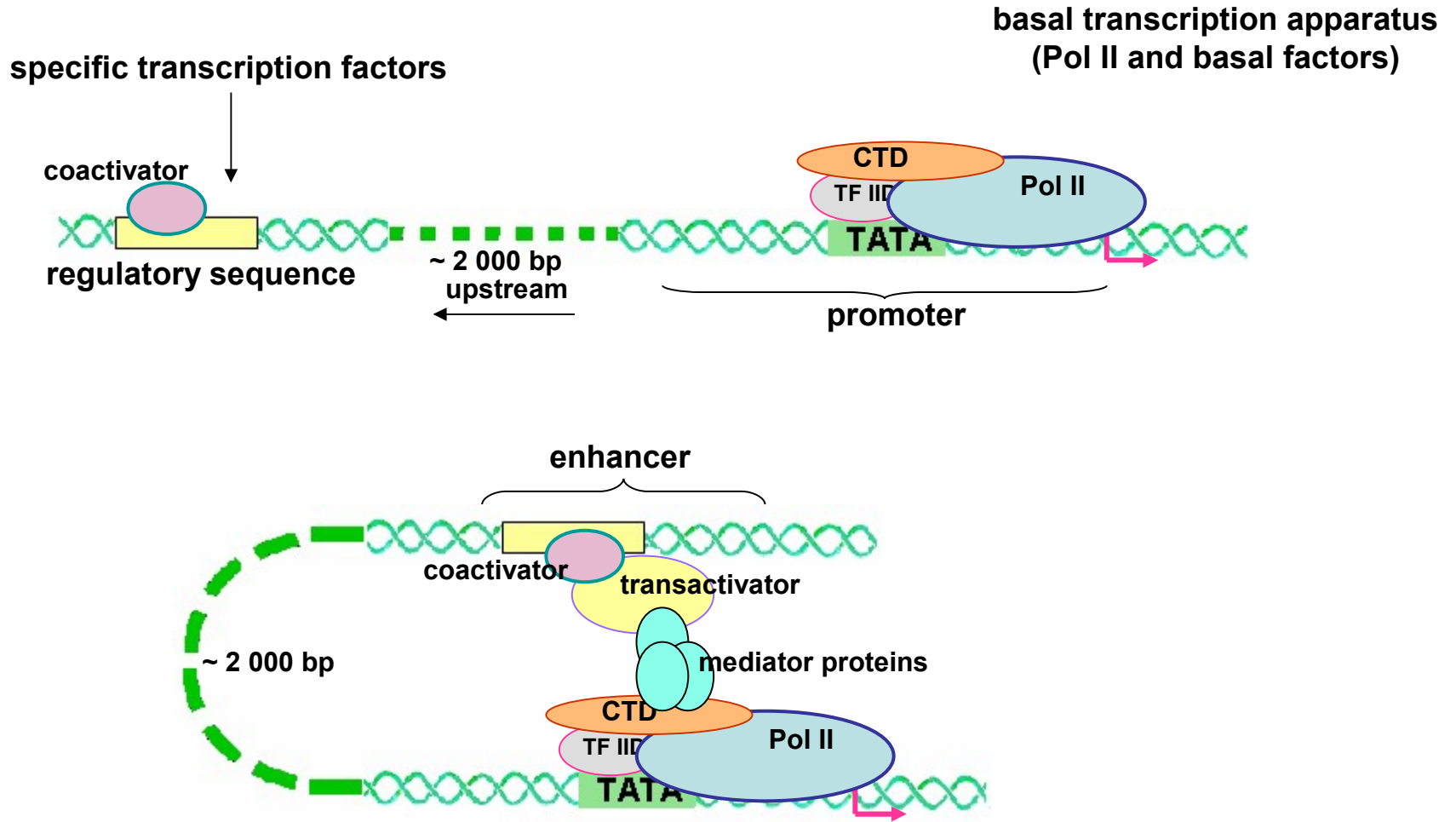
They mediate the effects of enhancers, silencers, and hormone response elements through interactions with other **mediator proteins** that interact directly with basal transcription factors and support or disable transcription of particular genes.

## Regulation of the function of transcription factors (both basal and specific)

- The synthesis of transcription factors (down- and up-regulation).
- The effects of transcription factors can be modulated by binding of stimulatory or inhibitory ligands, and also by cooperation of transcription factors.
- Factors can be phosphorylated or dephosphorylated owing to various extracellular signals (growth factors, peptidic hormones, cytokines, etc.).



# Regulation of a typical eukaryotic gene by an enhancer



## Regulation of transcription by steroid and thyroid hormones

Steroid and thyroid hormones (iodothyronines) are hydrophobic so that they can diffuse through the plasma membrane into the cells.

Hormones are bound onto **specific intracellular receptors**.

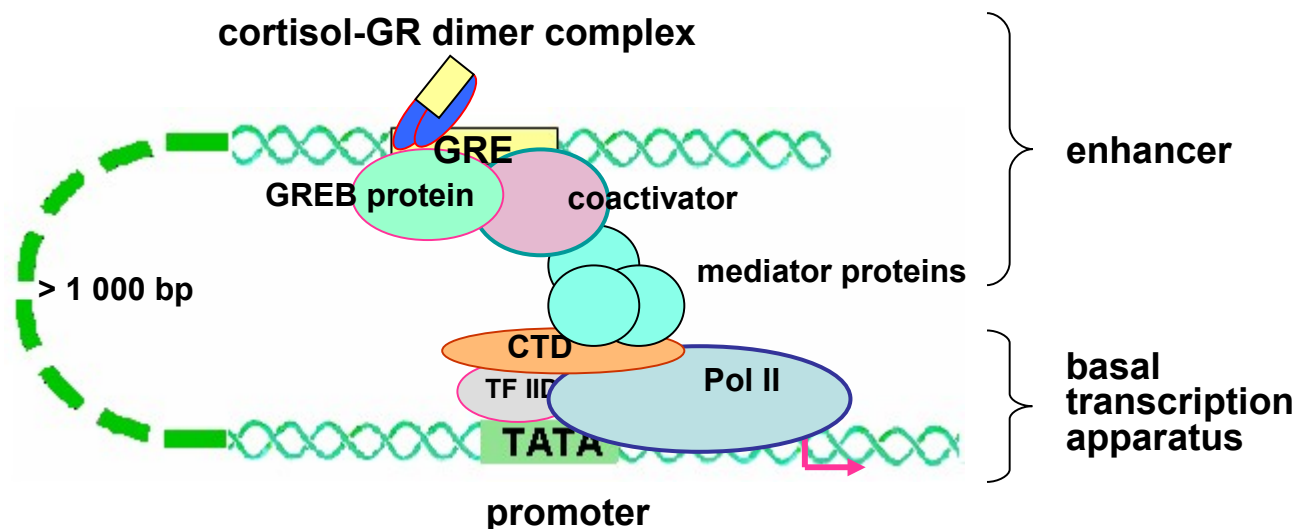
Complexes of these receptors with hormones are **specific transcription factors**. They bind onto regulatory DNA sequences called **hormone response elements (HRE)**.

The interaction with coactivators and mediator proteins follows and interaction between mediator proteins and the basal transcription apparatus initiates (or inhibits) the transcription of particular gene.

## Example: Initiation of transcription by cortisol

Active complex cortisol-receptor **binds onto DNA** at the specific sequence **GRE** (glucocorticoid response element, one of the HRE – hormone response elements).

The **coactivator** and specific **hormone response element-binding proteins** (HREB-proteins) are also attached. This complex acquires the ability to act as enhancer that supports initiation of transcription on the promoter by means of **mediator proteins**.



GR dimer – intracellular glucocorticoid receptor (dimer)

GRE – glucocorticoid response element

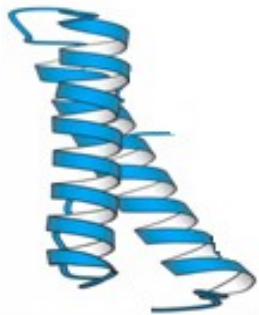
GREB protein – GRE binding protein (a specific transcription factor)

# Transcription factors that bind onto regulatory DNA sequences

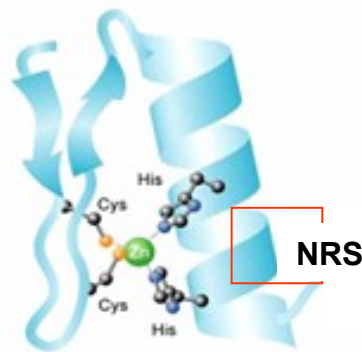
comprise mostly one of the **typical structural motifs**:

**helix-turn (or loop)-helix, zinc-finger, and leucine zipper.**

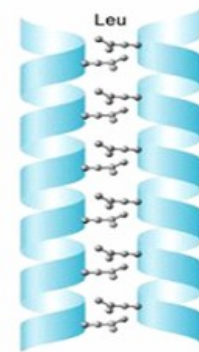
Only the small part of protein molecule (called DNA-binding domain) is responsible for the interaction with DNA. It is usually represented by two adjacent  $\alpha$ -helical segments.



helix-turn-helix



zinc finger



leucine zipper

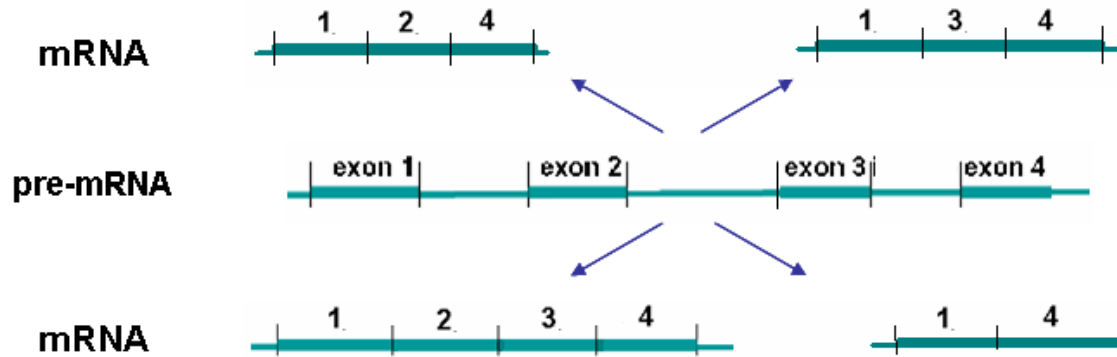
**Zinc finger**, e.g., occurs in DNA binding domains of steroid-hormone receptors. **NRS** (nucleotide recognition signal) is a part of  $\alpha$ -helix containing amino acid sequence that is able to recognize specific regulatory sequence of nucleotides in the major groove DNA.

Transcription factors are attached to DNA usually in the major groove.

# 3 Regulation at the level of processing of primary transcripts

## Alternative splicing

Alternative splicing can cause that the products of a sole gene are various proteins:



## RNA editing

In some mRNAs, the base sequence is altered after transcription by processes other than RNA splicing. Those processes are called **RNA editing** and are not very rare **E.g.**, cytidine residue may be deaminated to uridine, adenosine to inosine.

## 4 Regulation at the level of translation

is mediated mostly through changes in activities of eukaryotic initiation factors (eIFs).

### Example:

The synthesis of globin in reticulocytes is controlled by phosphorylation of the initiation factor eIF2. It is active when phosphorylated, inactive in the dephosphorylated form.

Haem prevents eIF2 from phosphorylation.

If haem is present within the cell, eIF2 is not phosphorylated - active, the translation of mRNA for globin chains proceeds.

If there is no haem in the cell, eIF2 is inactive and globin chains are not synthesized.