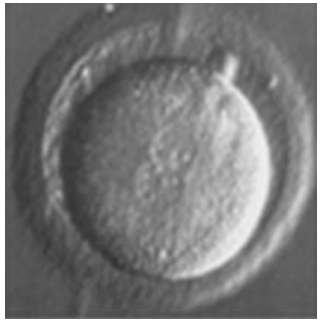
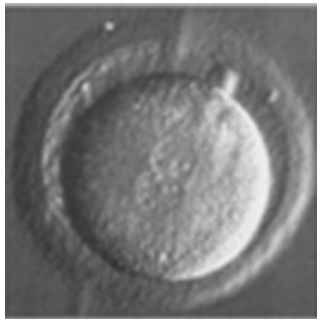
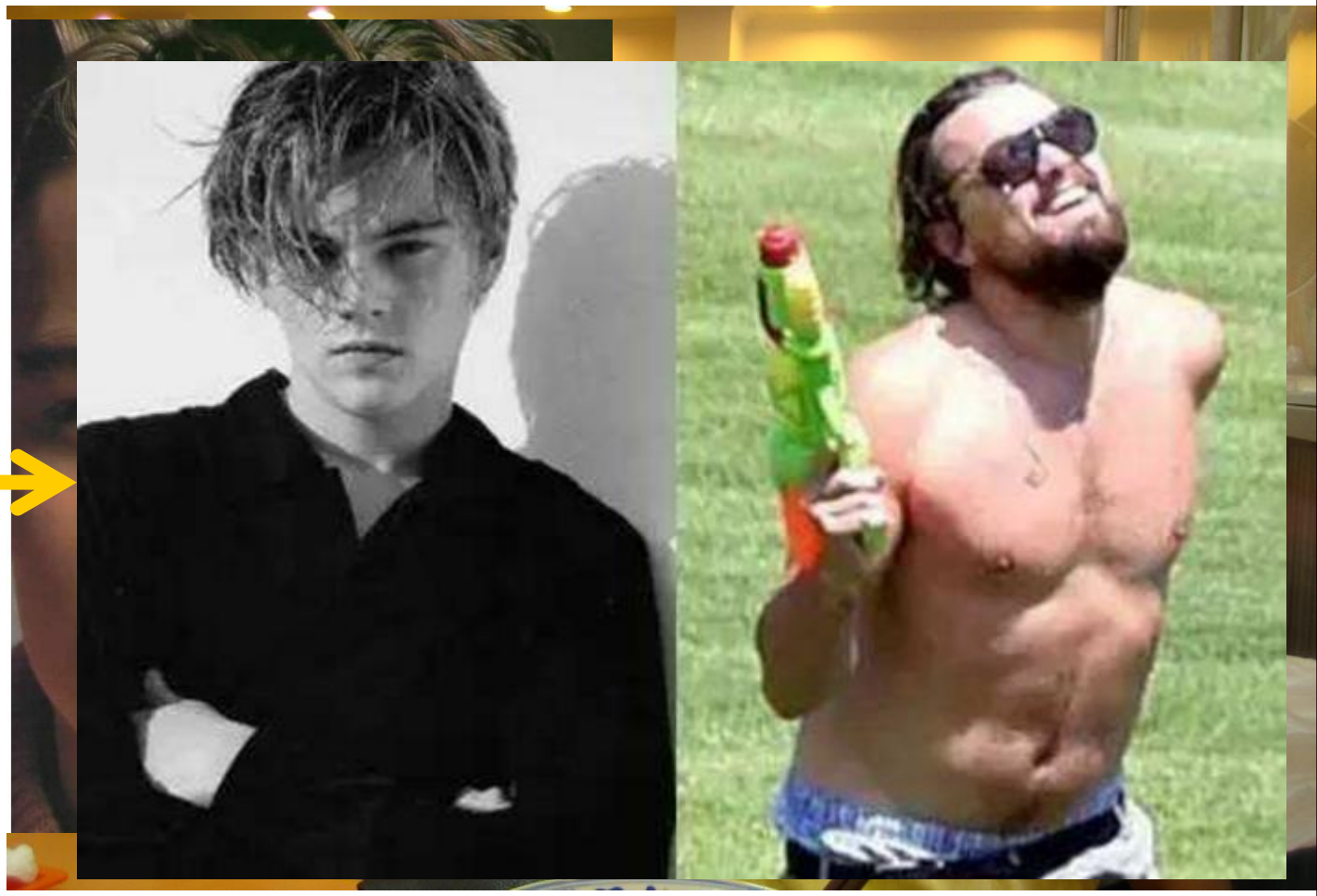
Regulation of gene expression

Same DNA→ complex organism
human embryo



Same DNA → diverse features

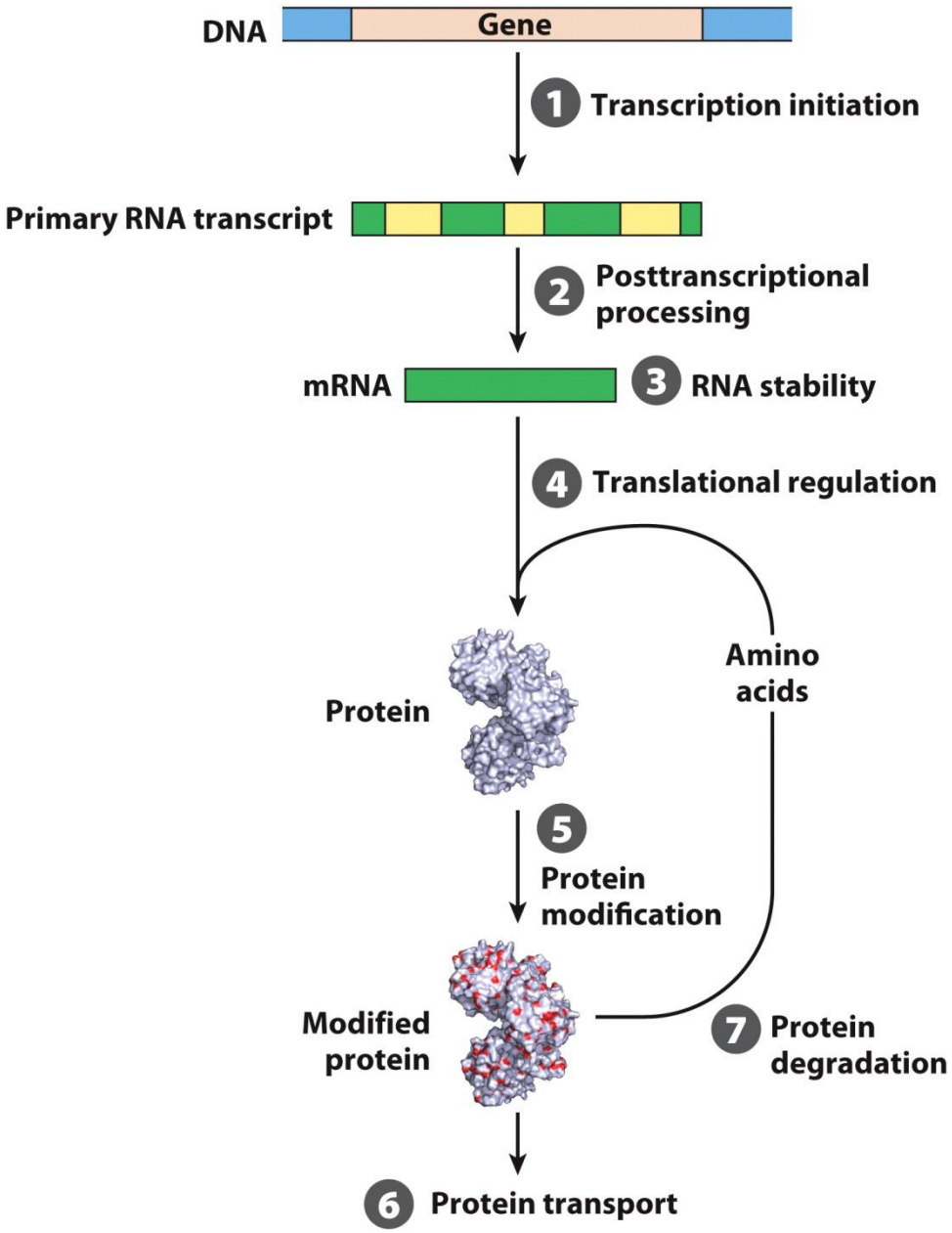




Highly similar DNA

Human & Chimpanzee DNA:
96% identical

Regulation of protein level in cell

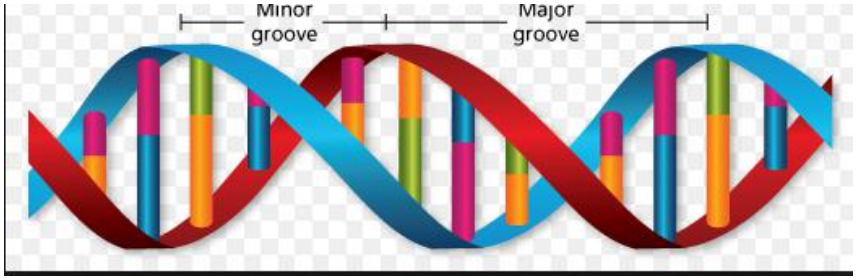


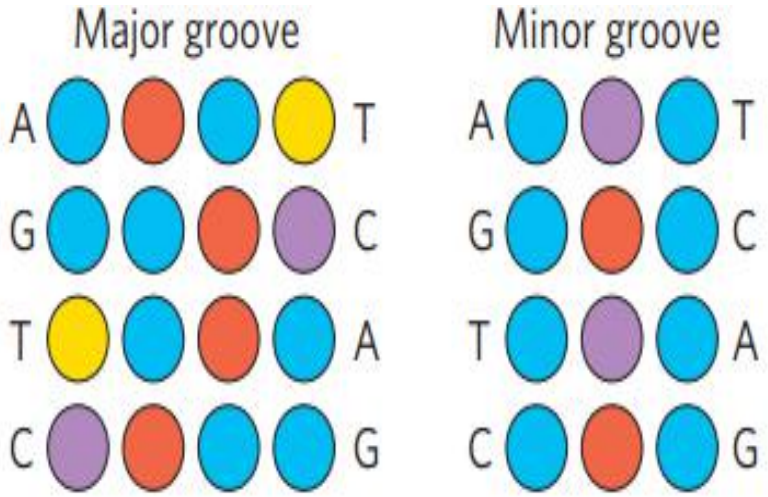
1. Synthesis of the primary RNA transcript (transcription)
2. Posttranscriptional modification of mRNA
3. Messenger RNA degradation
4. Protein synthesis (translation)
5. Posttranslational modification of proteins
6. Protein targeting and transport
7. Protein degradation

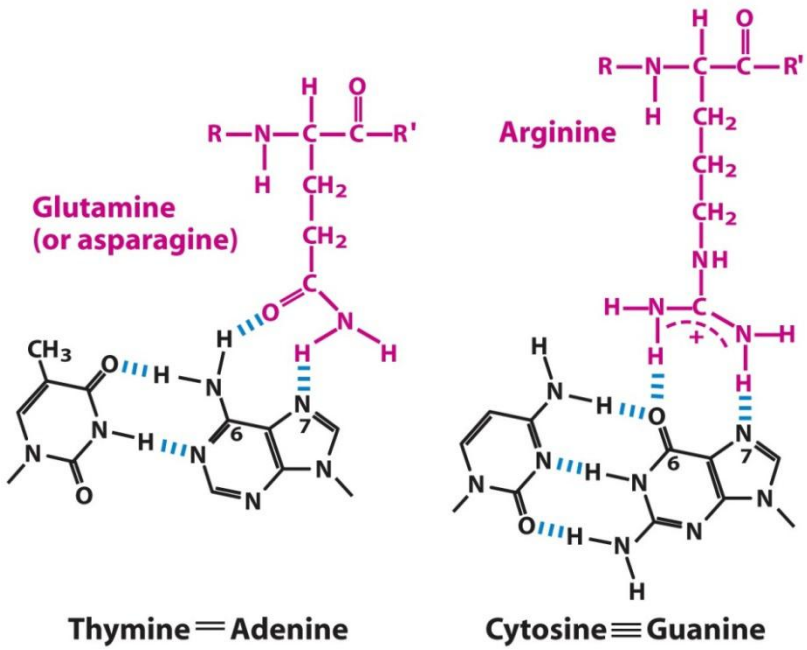
Principles of gene regulation

- RNA polymerase binds to DNA at promoters
- Transcription initiation is regulated by proteins that bind to or near promoters
- Regulatory proteins have discrete DNA-binding domains:
  helix-turn-helix, zinc finger and homeodomain, etc.
- Regulatory proteins also have protein-protein interaction domains:
  Leucine zipper and basic helix-loop-helix, etc.

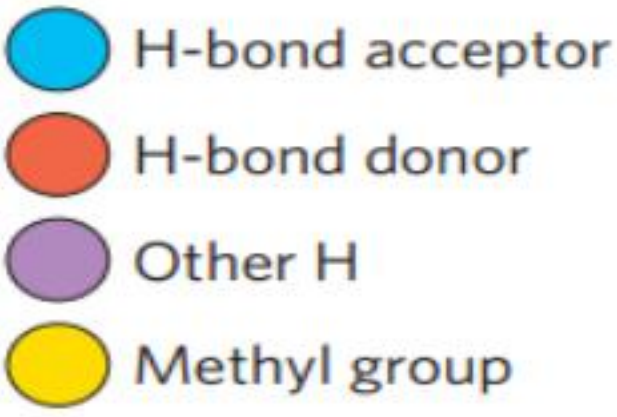
DNA-protein binding in transcription regulation







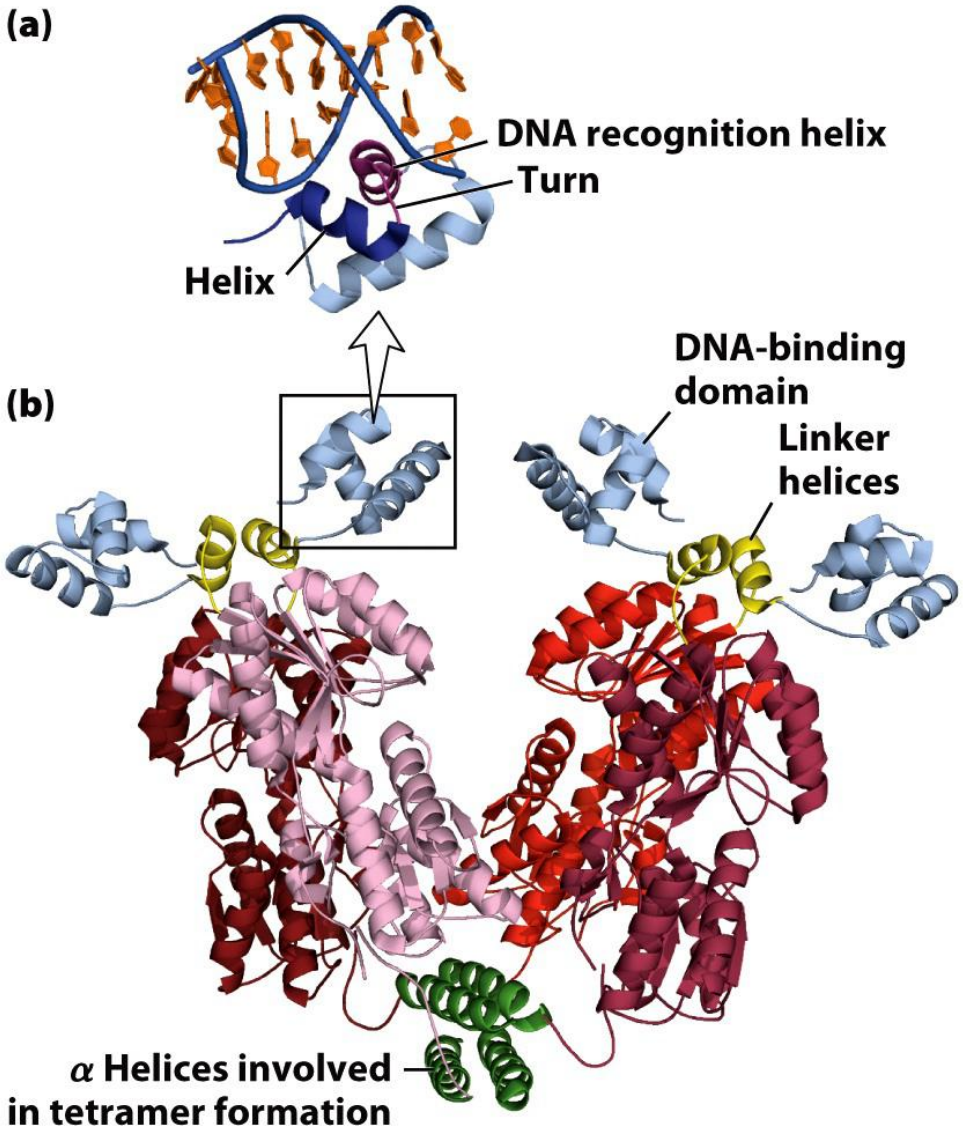
Two examples of specific amino acid–base pair interactions



To bind specifically to DNA sequences, regulatory proteins must recognize surface features on the DNA. Most of the chemical groups that differ among the four bases and thus permit discrimination between base pairs are hydrogen-bond donor and acceptor groups exposed in the major groove of DNA.

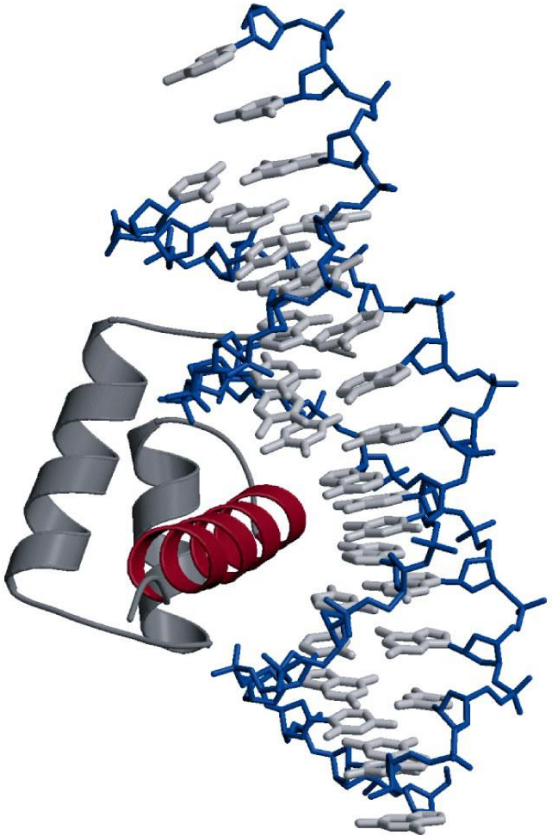
Helix-turn-helix

- The helix-turn-helix motif comprises about 20 amino acids in two short α-helical segments, each 7 to 9 amino acid residues long, separated by a β-turn.
- The helix-turn-helix DNA-binding motif is crucial to the interaction of many regulatory proteins with DNA.



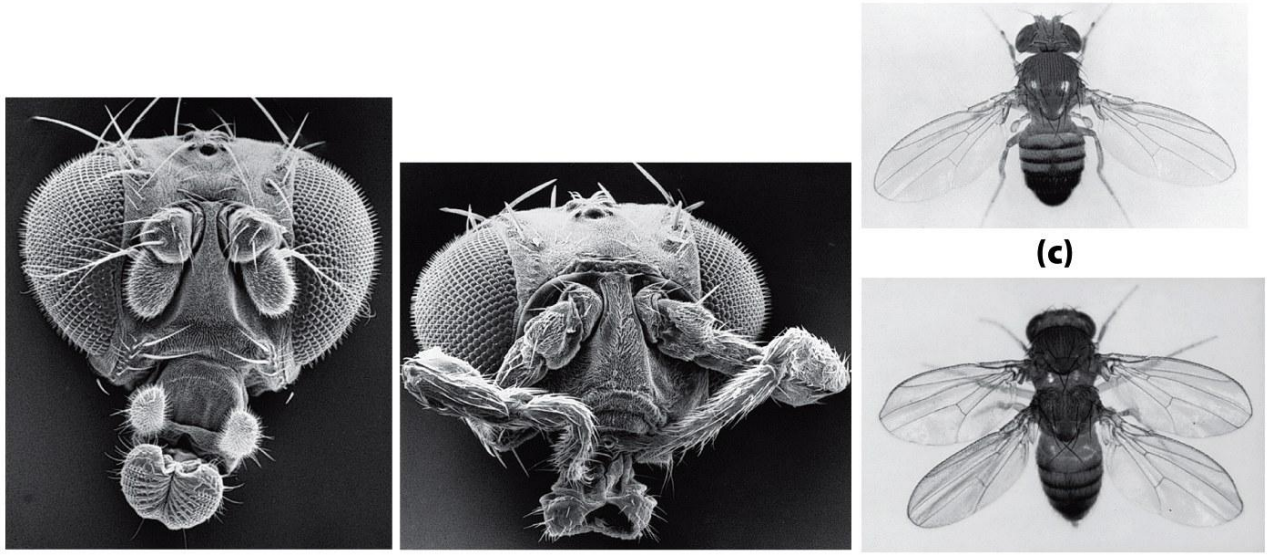
Lac repressor

Homeodomain, found in Homeotic genes that determine which parts of the body form what body parts



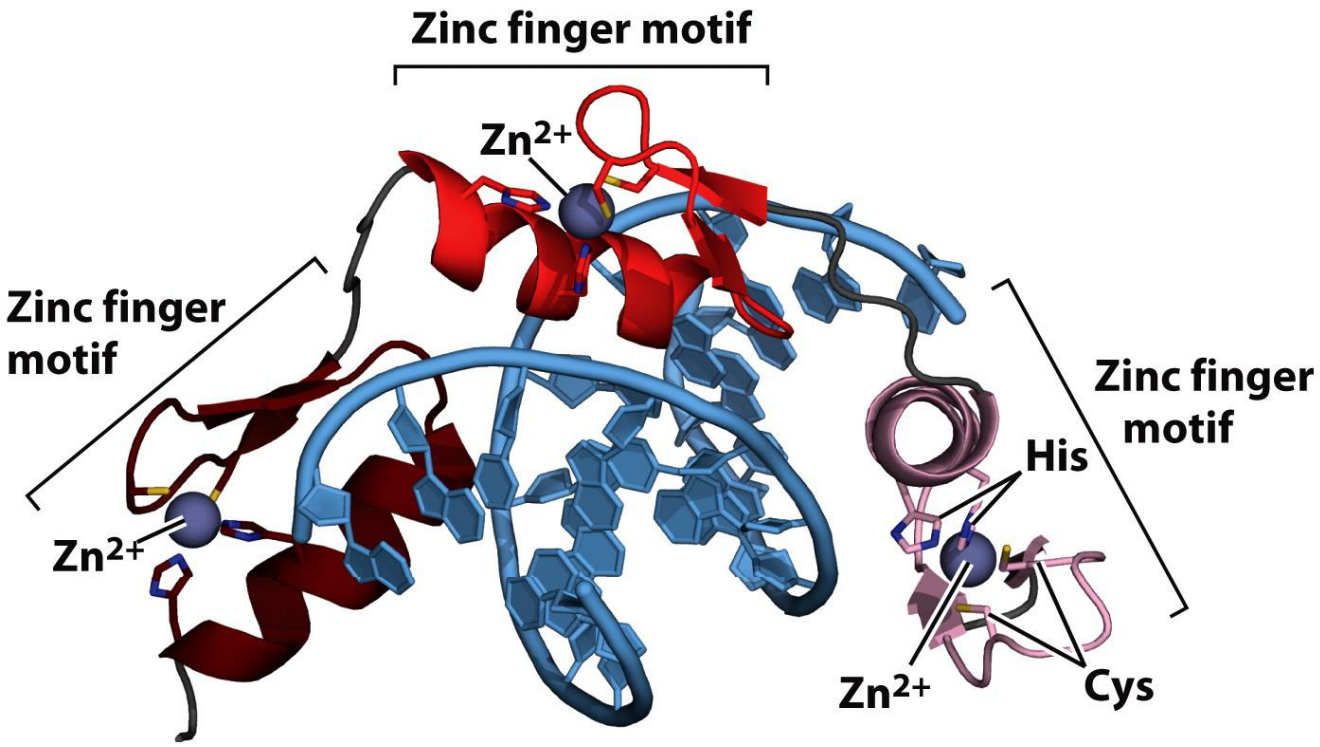
Homeodomain in protein Ultrabithorax (ubx)

- **Homeobox** is the DNA sequence that encodes **Homeodomain**
- **Homeodomain** is highly conserved in a wide variety of organisms, including humans

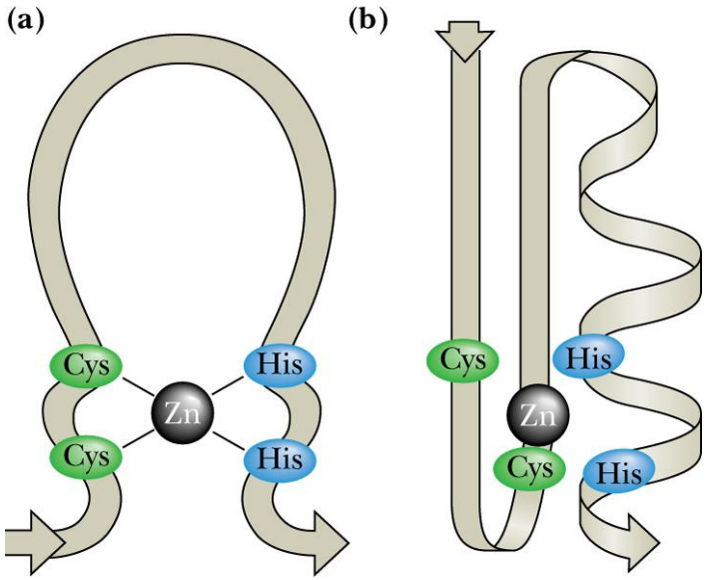


Effects of mutations in Hox (homeobox) genes in *Drosophila.*
Loss of homeotic genes by mutation or deletion causes the appearance of a normal appendage or body structure at an inappropriate body position.

Zinc finger



3 zinc fingers in regulatory protein Zif68



Binding of the Zn^{2+} ion by four Cys (or two Cys and two His) residues stabilizes an elongated loop structural motif that is important for DNA binding. The zinc itself does not interact with DNA.

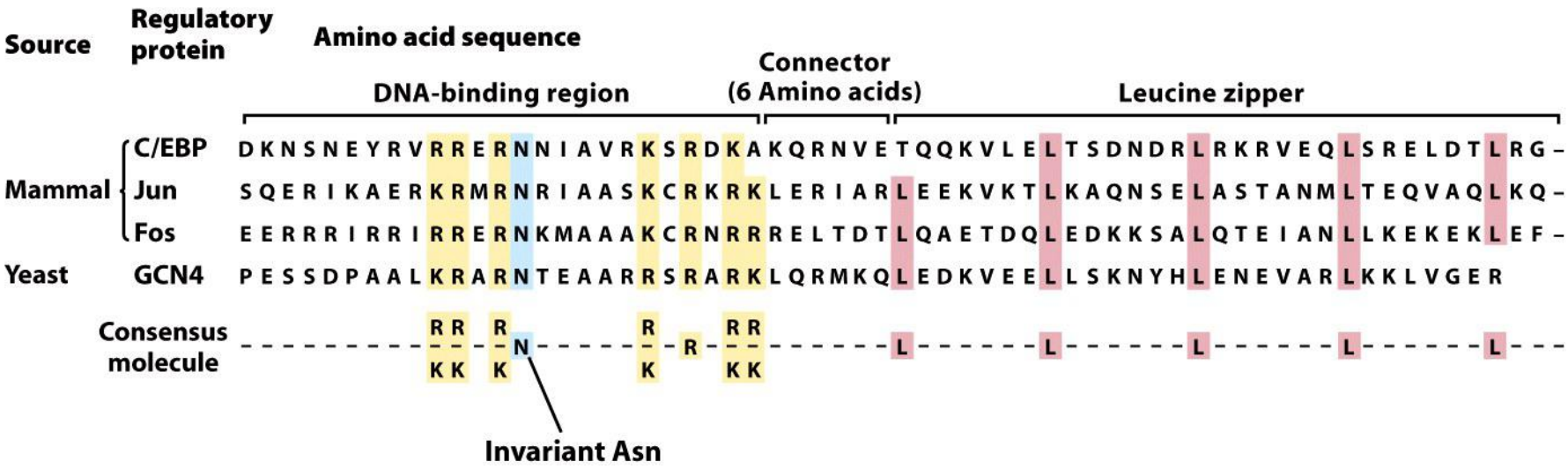
Protein-protein interaction in regulation of transcription

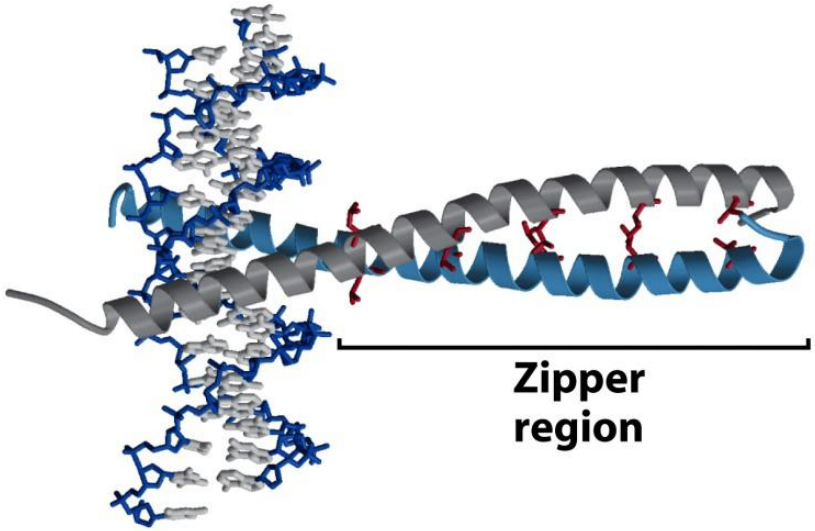
- Regulatory proteins contain domains not only for DNA binding but also for protein-protein interactions

- Protein-Protein interaction domains:

  Leucine zipper, basic helix-loop-helix, & much more others

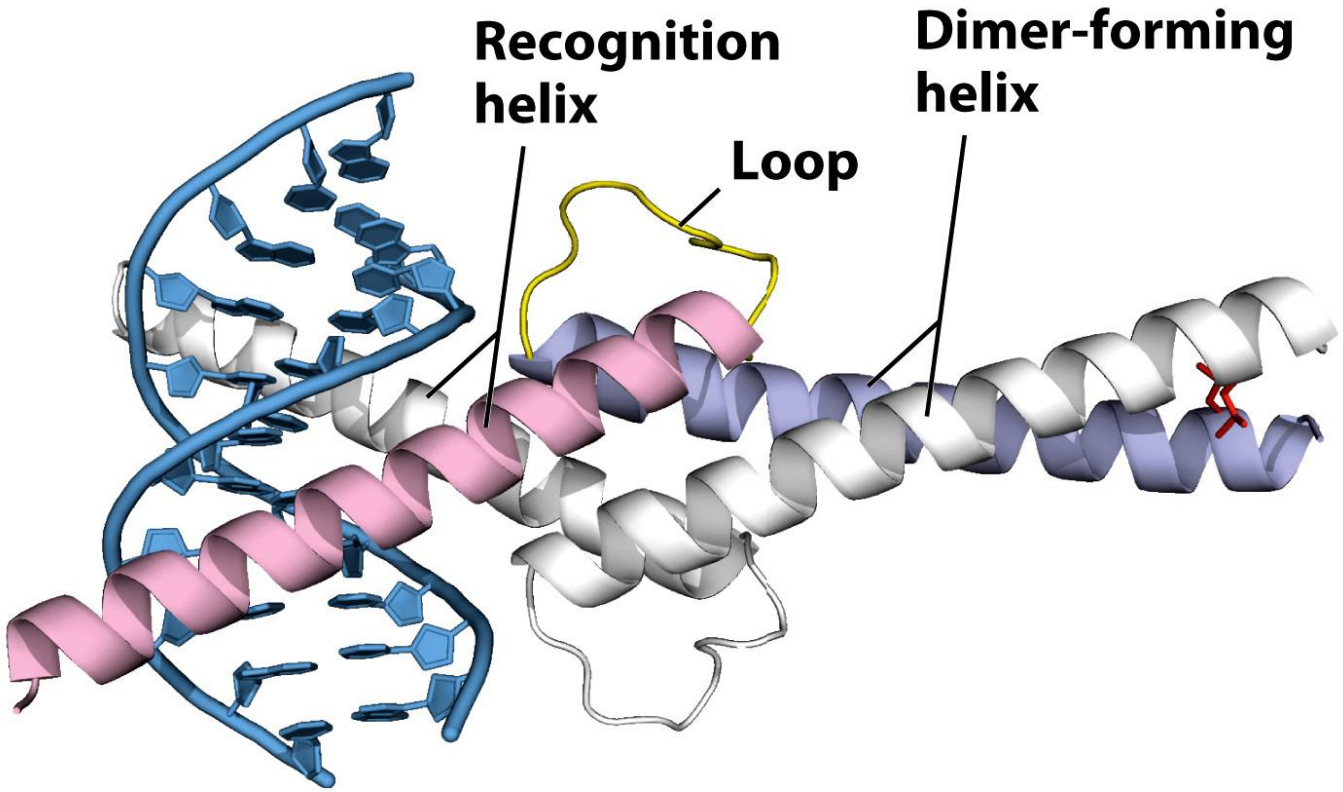
Leucine Zipper





- **Leucine zipper** is an amphipathic α helix with hydrophobic amino acids concentrated on one side forming contact surface between the two polypeptides of a dimer. These α helices have Leu residues at every 7th position, forming a straight line along the hydrophobic surface.
- Regulatory proteins with leucine zippers often have a separate DNA-binding domain with a high concentration of basic (Lys or Arg) residues that can interact with the negatively charged phosphates of the DNA backbone.

Basic Helix-Loop-Helix



- **Basic Helix-Loop-Helices** have a conserved region of about 50 amino acid residues that is important in both DNA binding and protein dimerization.
- This region can form two short amphipathic α helices linked by a loop of variable length, the helix-loop-helix (distinct from the helix-turn-helix motif associated with DNA binding).

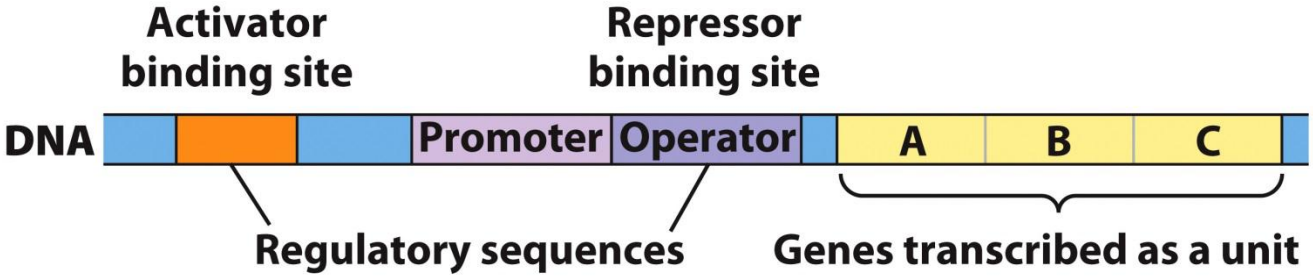
Regulation of gene expression

Constitutive gene expression: housekeeping genes

Regulated gene expression: other genes whose level change in response to cell signal (inducible & repressible)

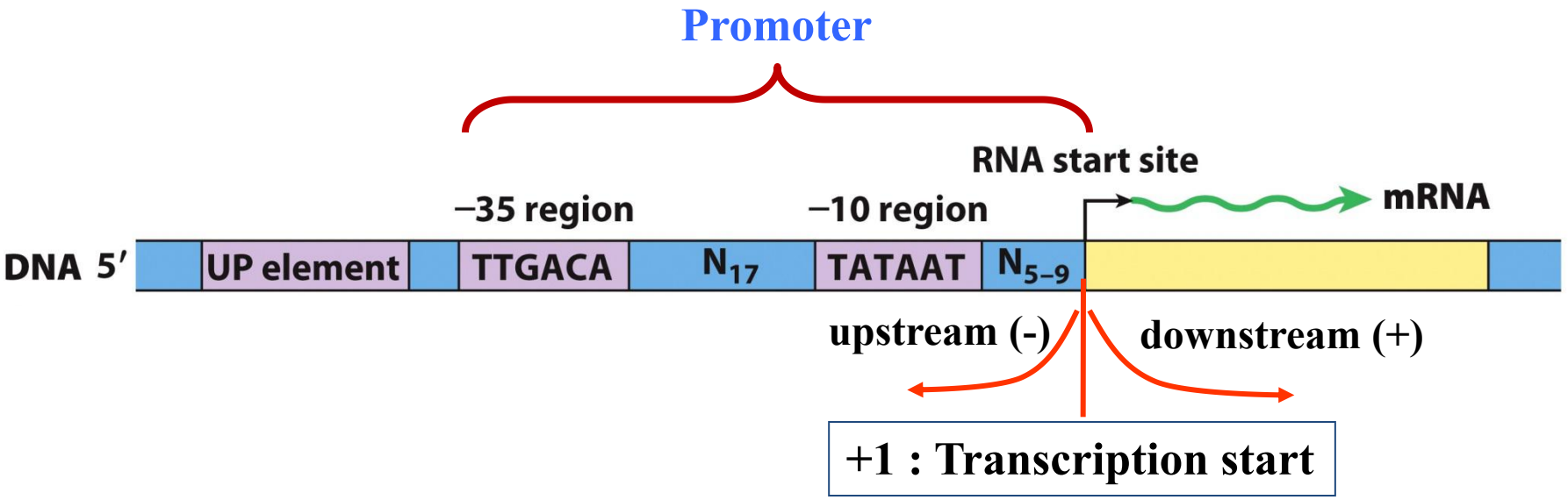
Operator (操纵基因): a region of DNA that interacts with a regulatory protein to control the expression of a gene or group of genes.

Operon (操纵子): in prokaryotes, a cluster of related genes with a promoter, plus additional sequences that function together in regulation of gene expression form a single transcript **(Polycistronic mRNA)**. The single promoter and operator can regulates the expression of all genes in the cluster.



The general organization of operons

RNA polymerase binds to DNA at promoters



Promoters: specific sequences in the DNA that the RNA polymerase holoenzyme can bind and direct the transcription of adjacent segments of DNA (genes).

Different σ subunits can recognize different promoters.

Regulation of transcription initiation:

transcription initiation is regulated by proteins that bind to or near promoters

At least three types of proteins regulate transcription initiation

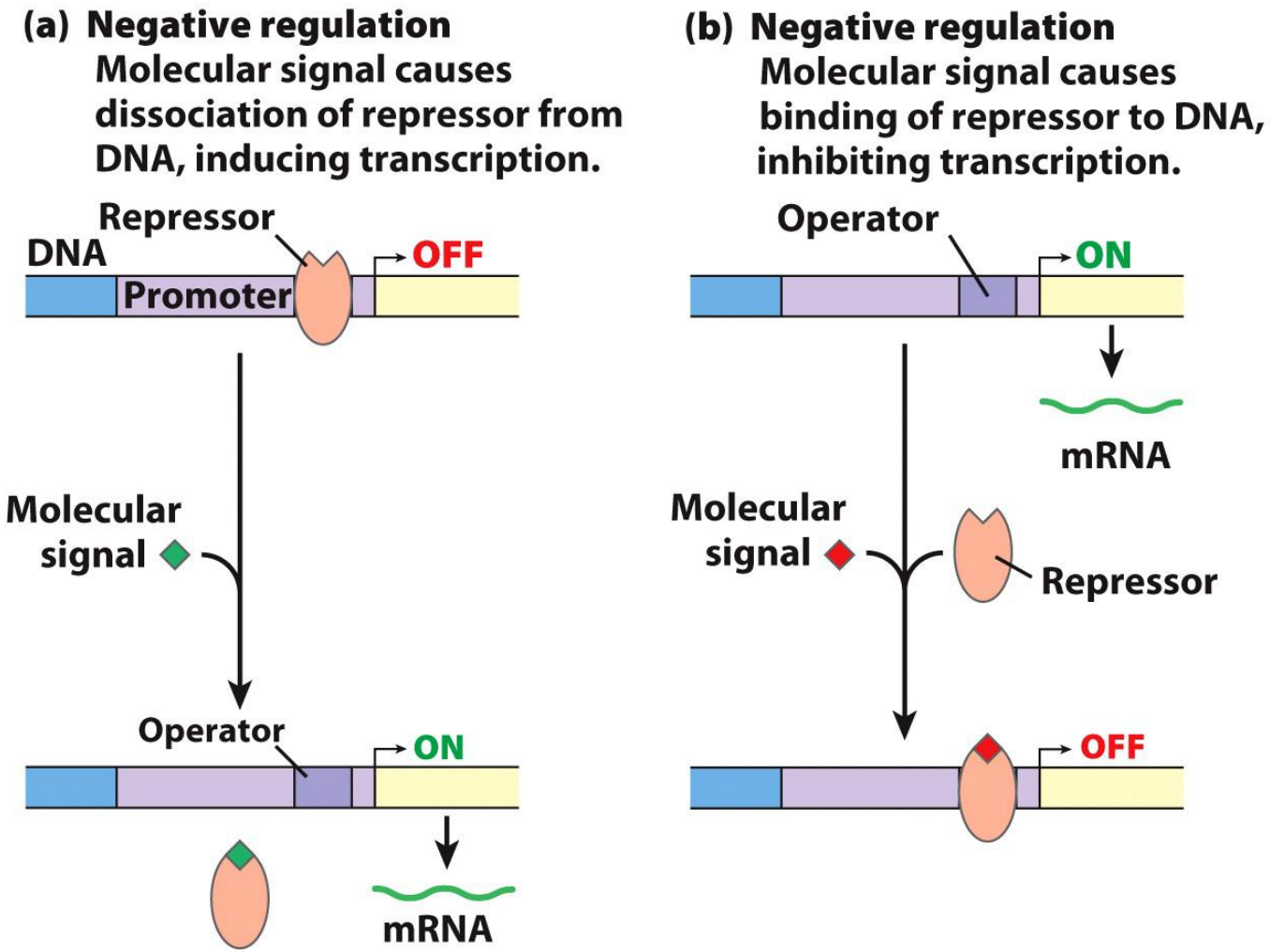
Specificity factors: alter the specificity of RNA polymerase for a given promoter or set of promoters;

Repressors: impede access of RNA polymerase to the promoter;

Activators: enhance the RNA polymerase–promoter interaction.

Patterns of regulation of transcription initiation

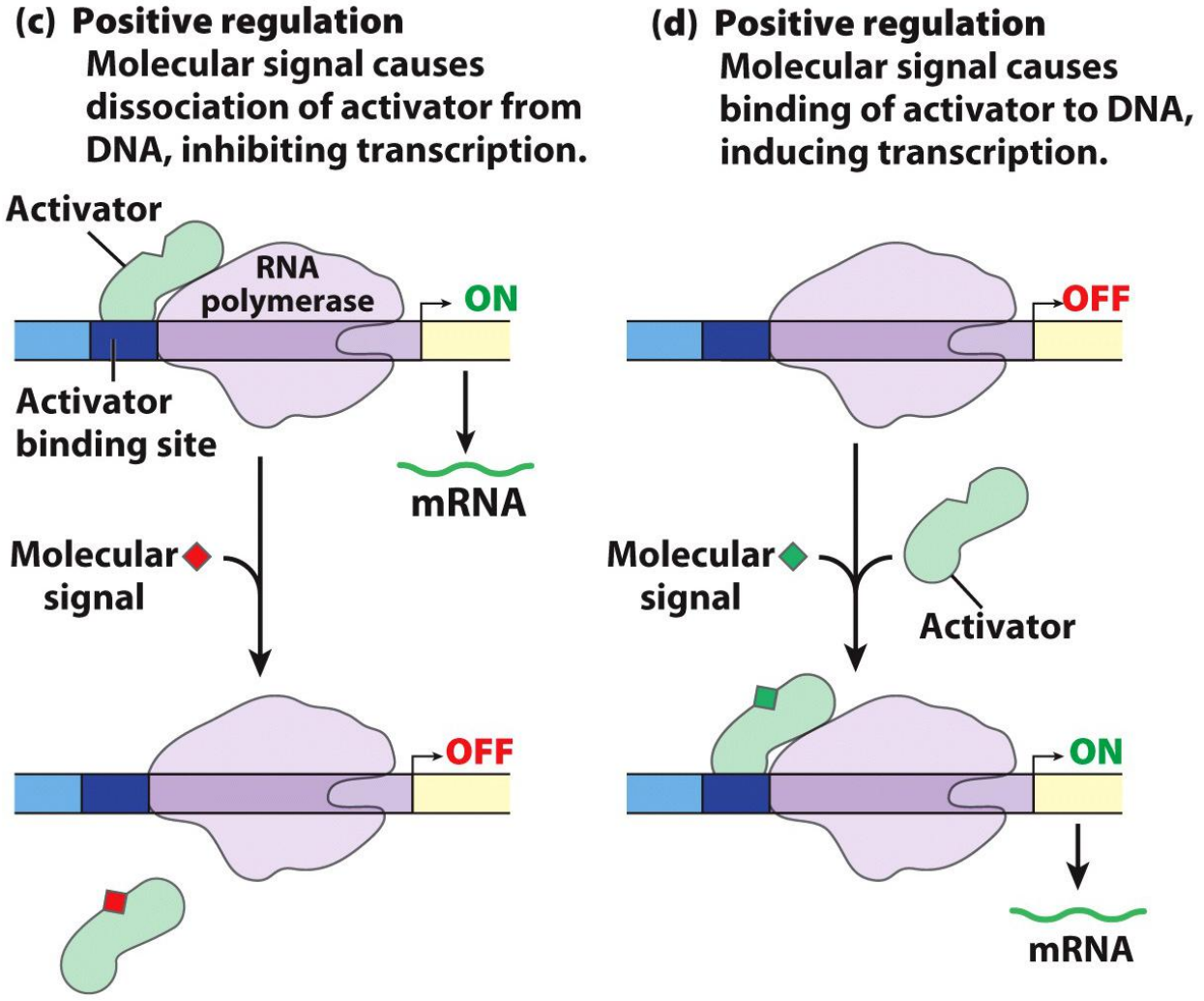
Negative regulation



Repressors: proteins that bind to the regulatory sequence (operator) for a gene, *block* its transcription.

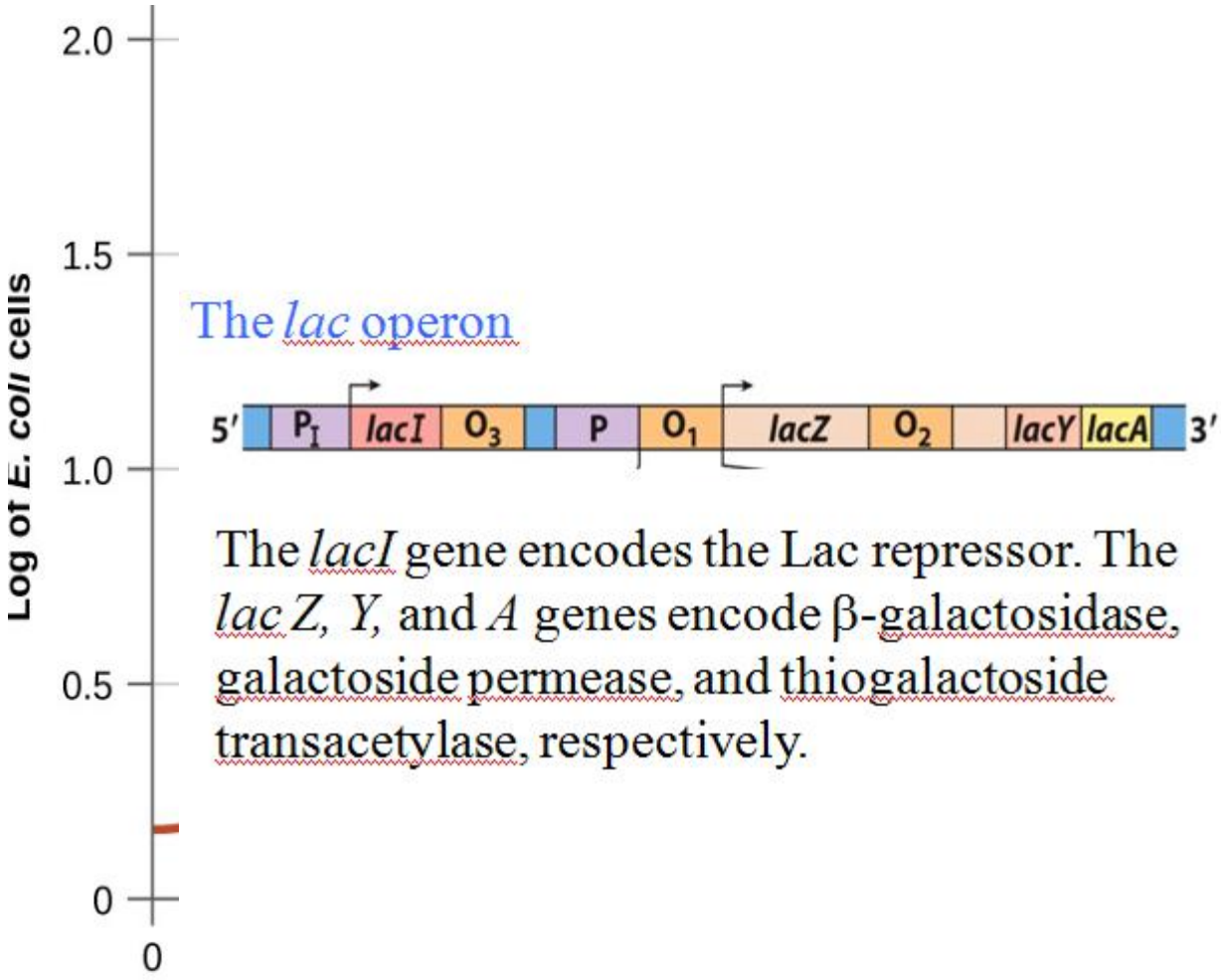
Patterns of regulation of transcription initiation

Positive regulation



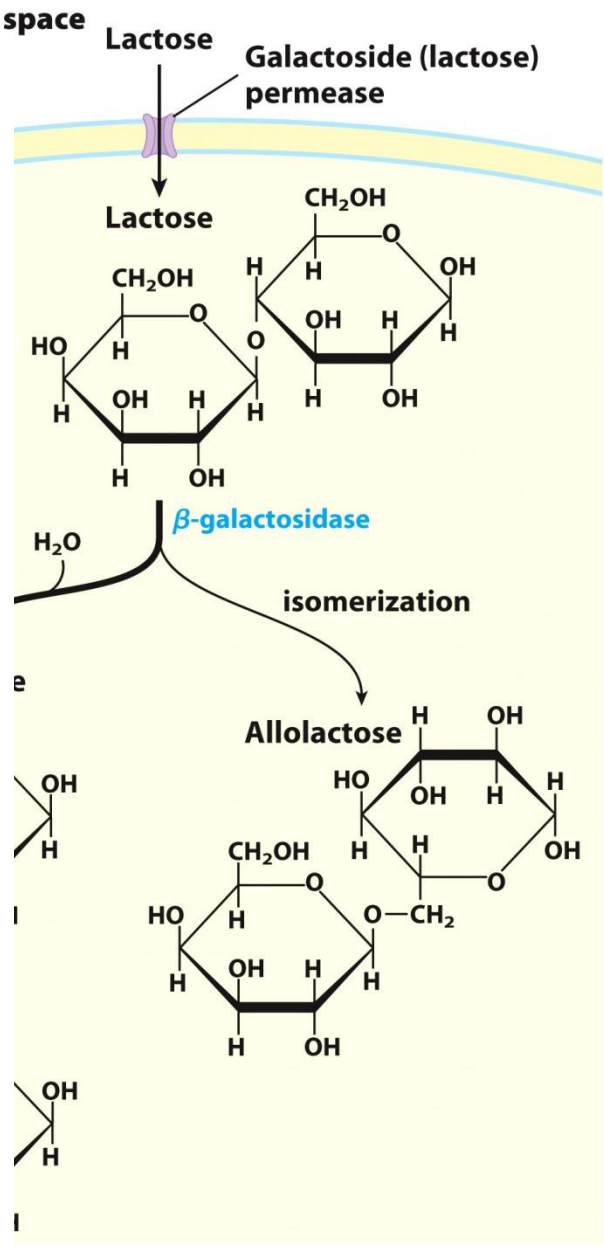
Activators: proteins that bind to DNA and *enhance* the activity of RNA polymerase at a promoter.

Lactose metabolism in *E.coli*

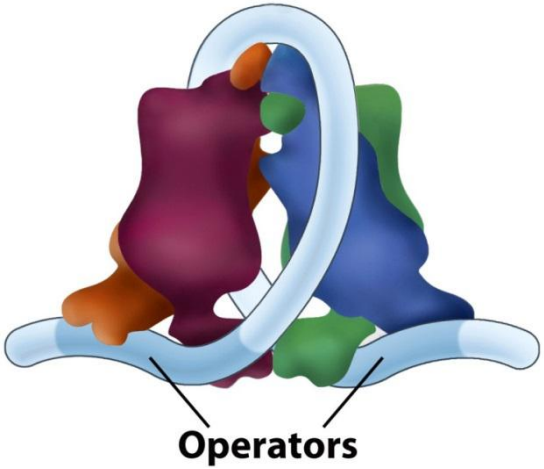


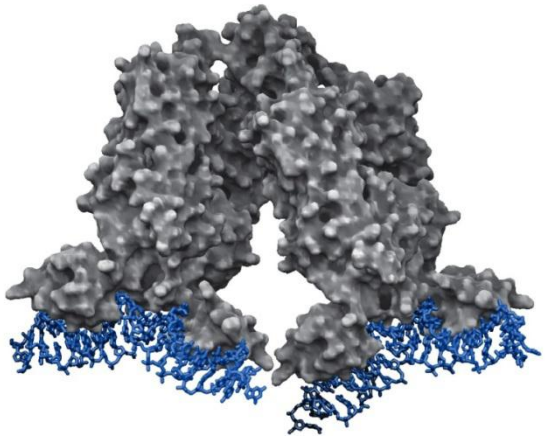
The *lac* operon

The *lacI* gene encodes the Lac repressor. The *lac Z, Y,* and *A* genes encode β-galactosidase, galactoside permease, and thiogalactoside transacetylase, respectively.

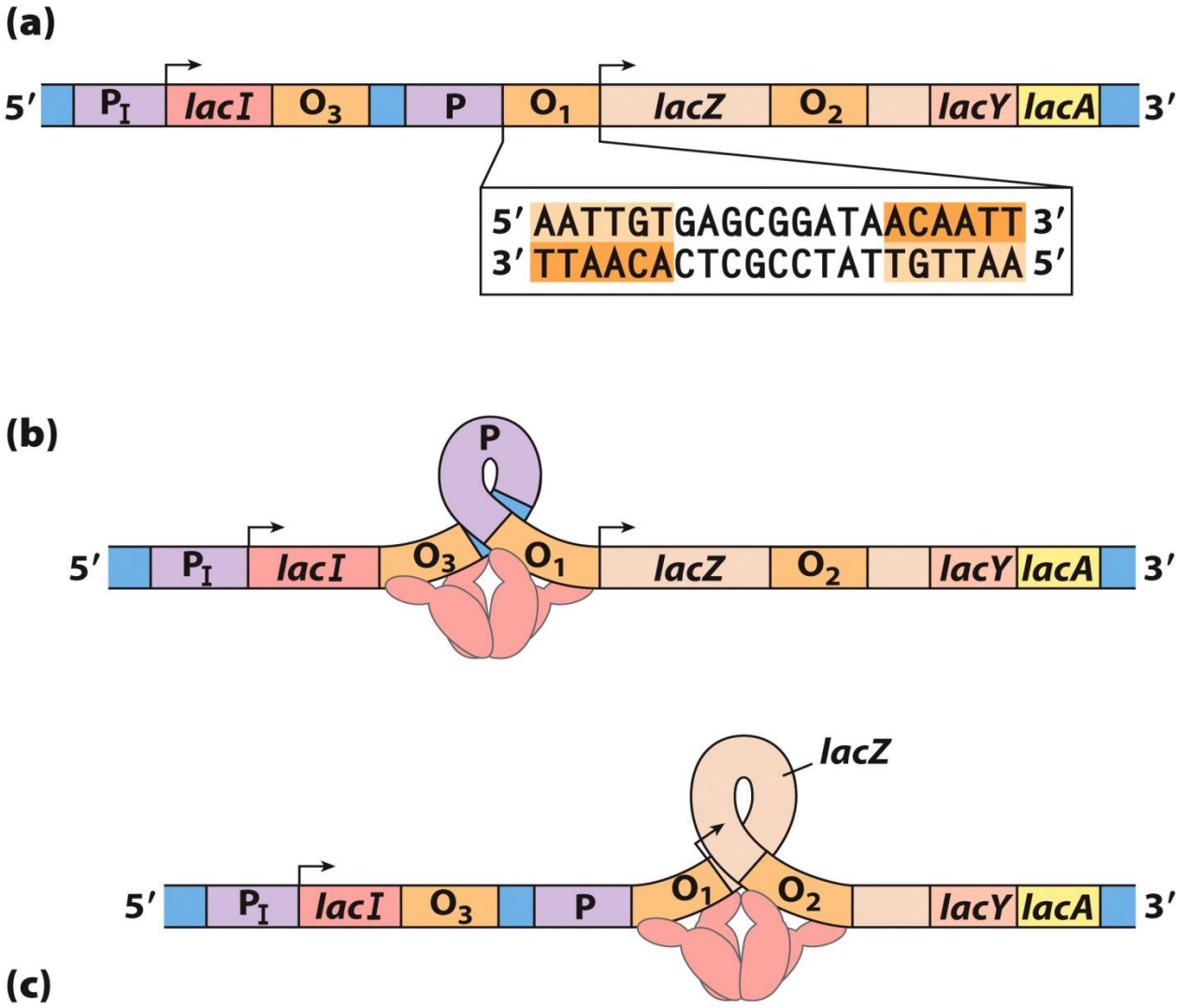


The *lac* operon





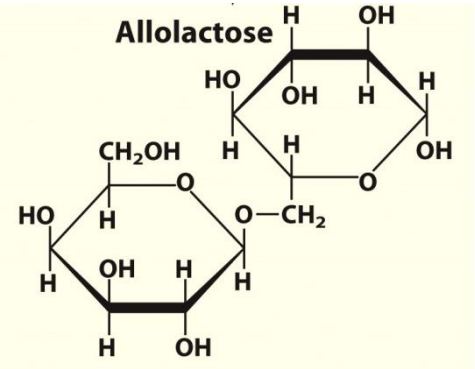
The Lac repressor binds to the operators



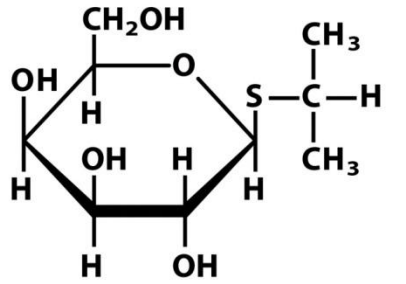
The *lac Z, Y,* and *A* genes encode β-galactosidase, galactoside permease, and thiogalactoside transacetylase, respectively.

Lac operon inducer

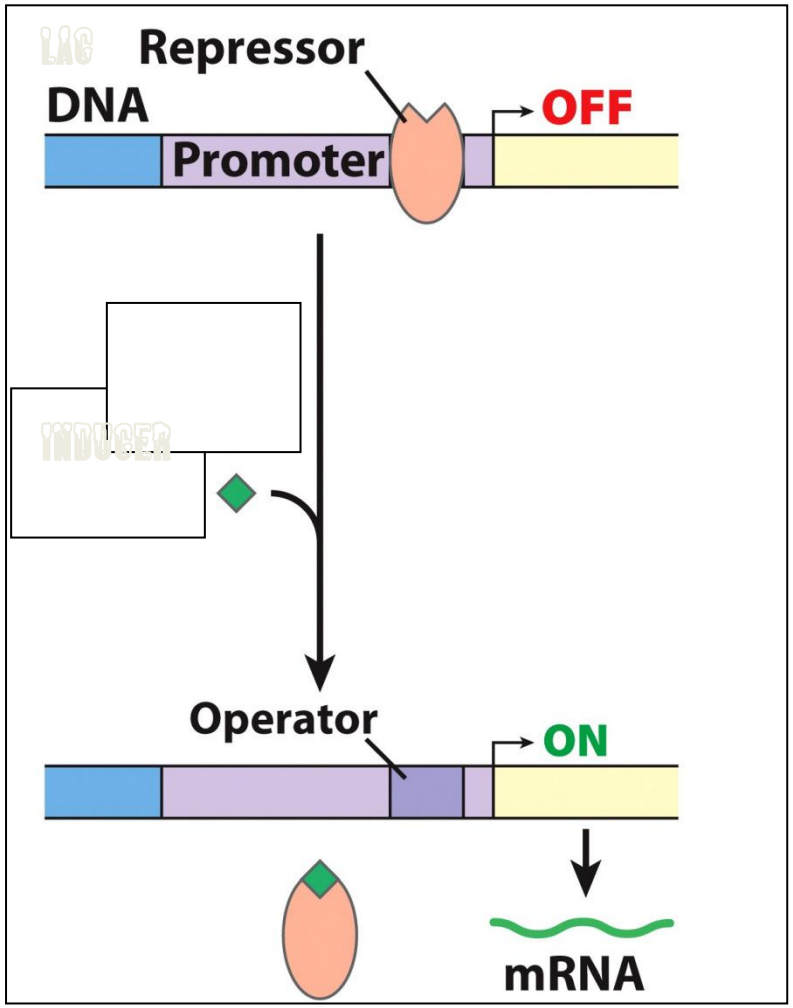
Inducer: A signal molecule that, when bound to a regulatory protein, produces an increase in the expression of a given gene.



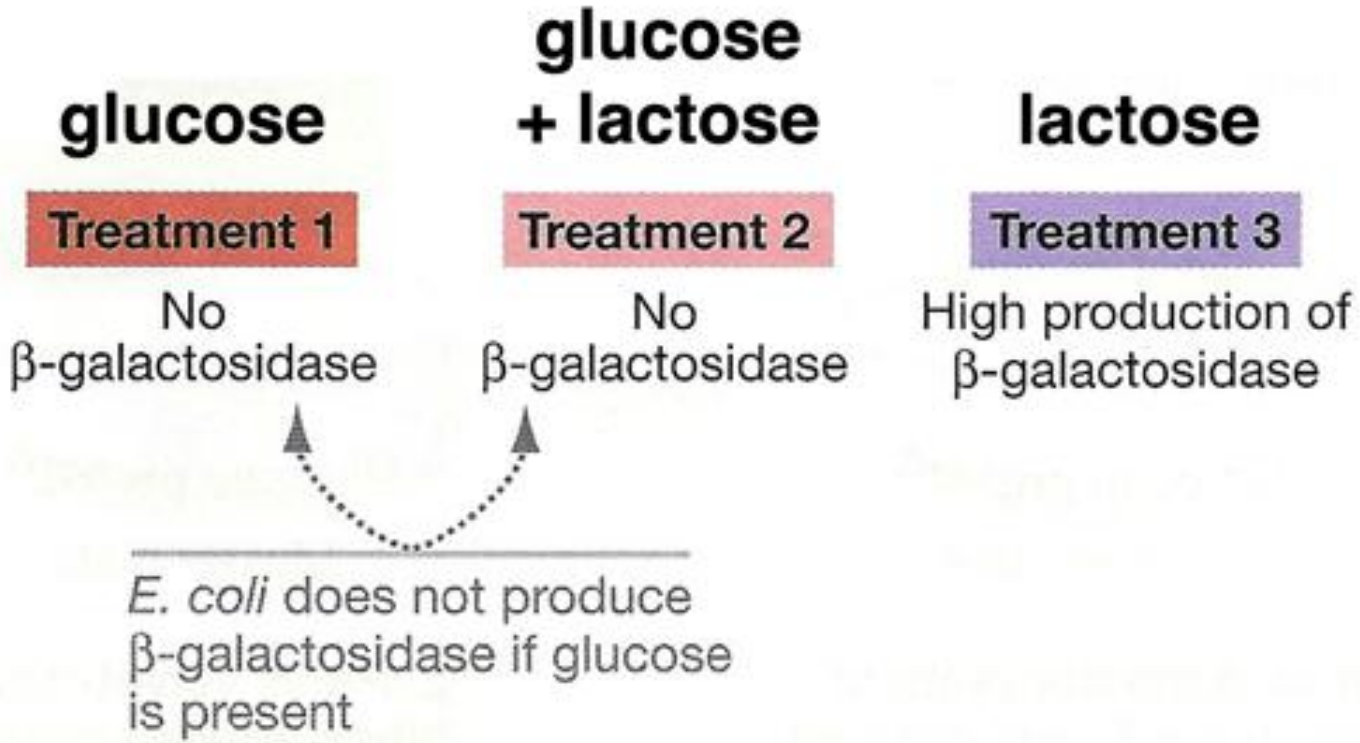
Allolactose: the inducer in the *lac* operon is not lactose itself but allolactose, an isomer of lactose.



IPTG: experimentally used inducer for the *lac* operon

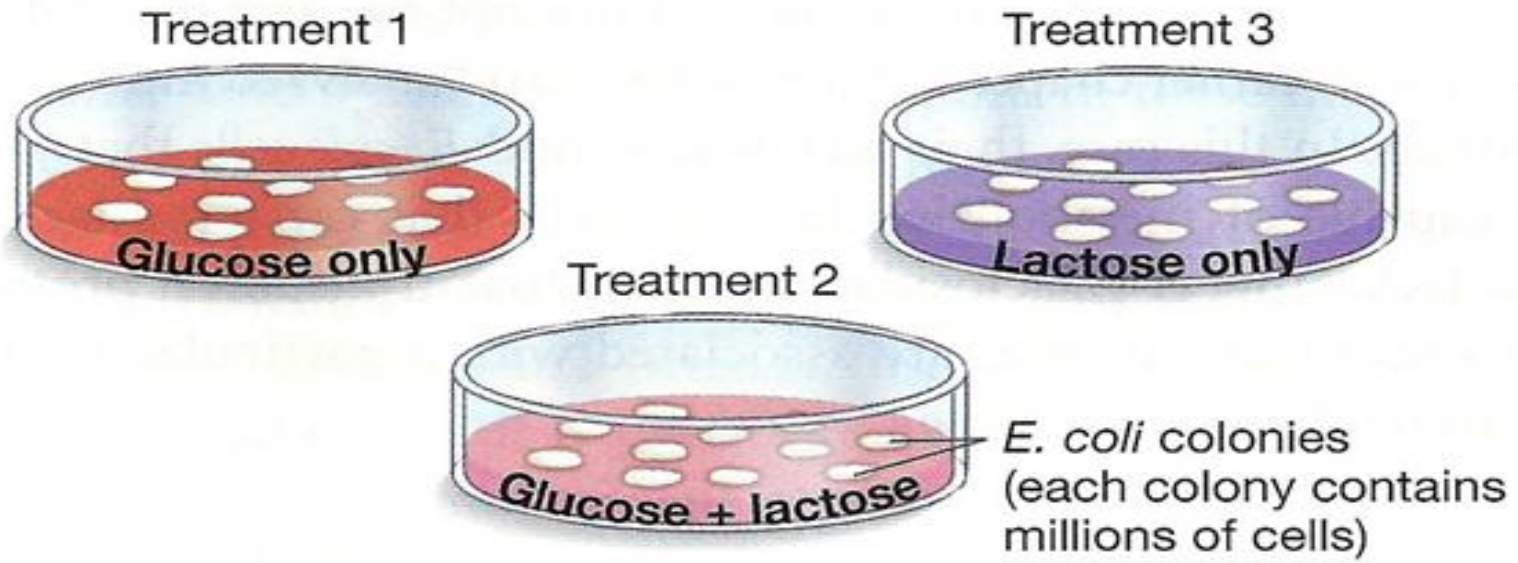


How does *E. coli* respond to lactose?



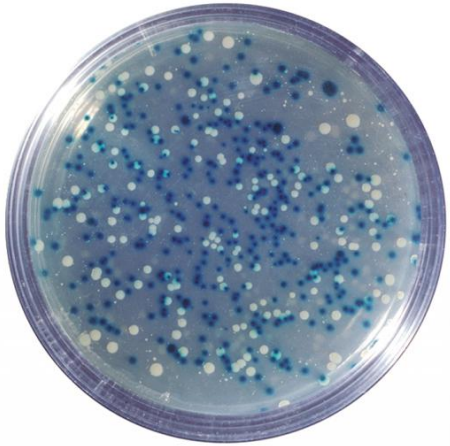
CONCLUSION: Glucose prevents expression of the gene for β-galactosidase. The presence of lactose without glucose stimulates expression of that gene.

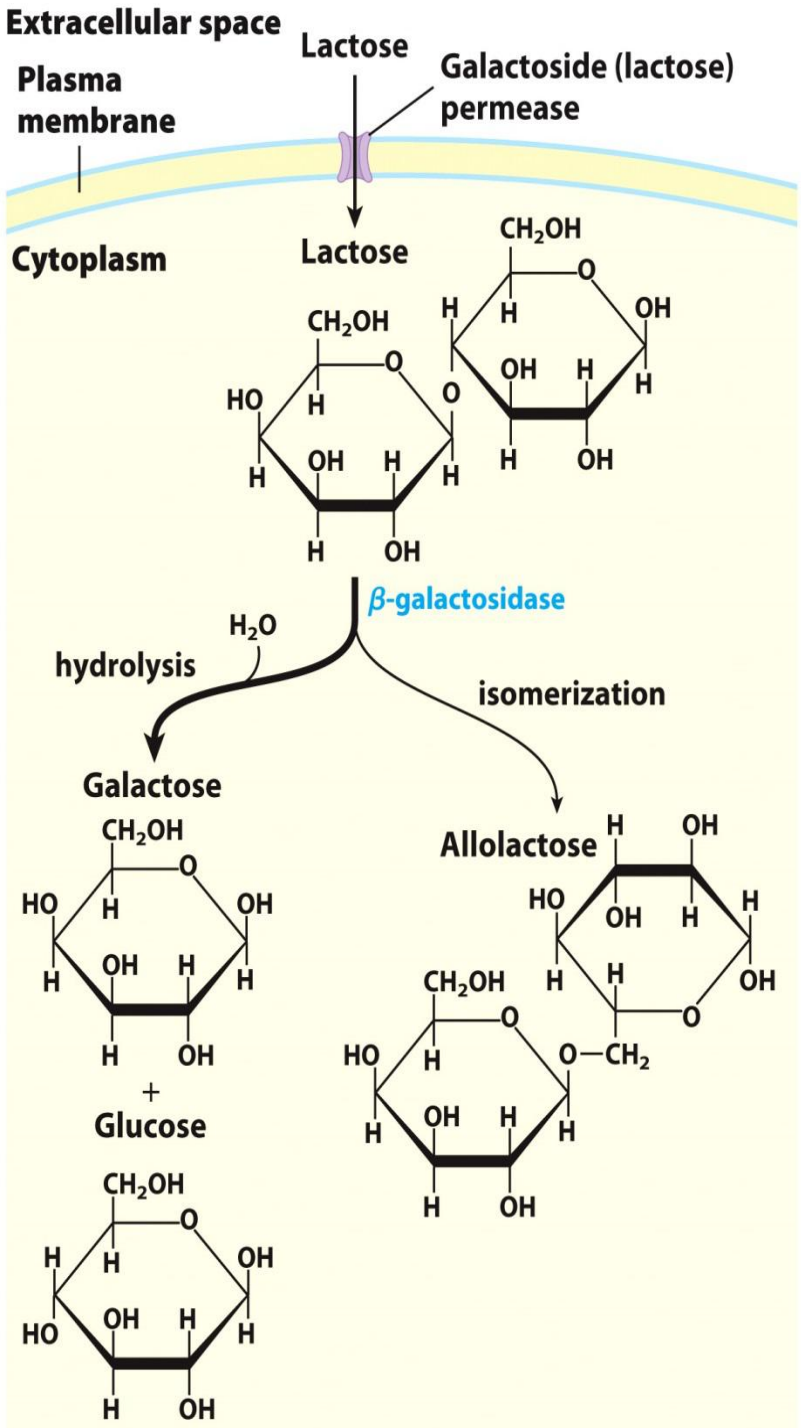
EXPERIMENTAL SETUP:



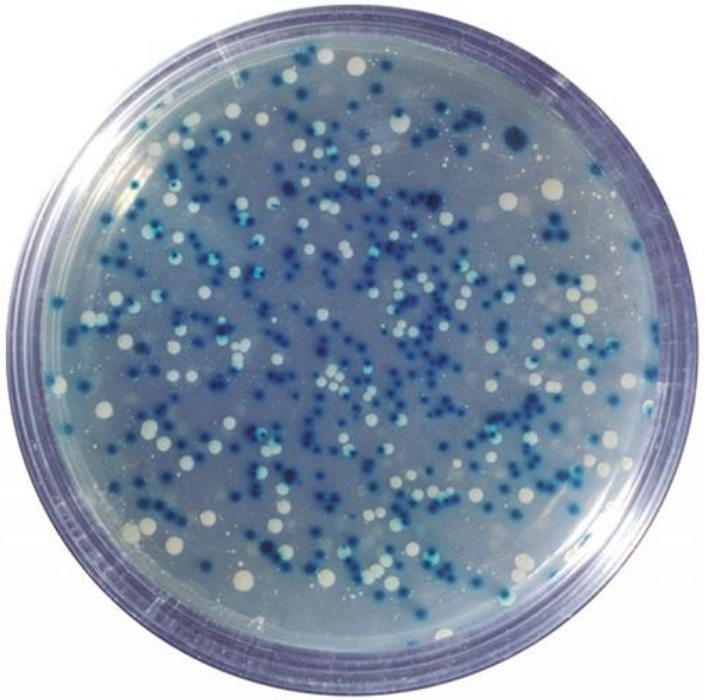
X-gal staining

X-gal identifies cells with beta-galactosidase

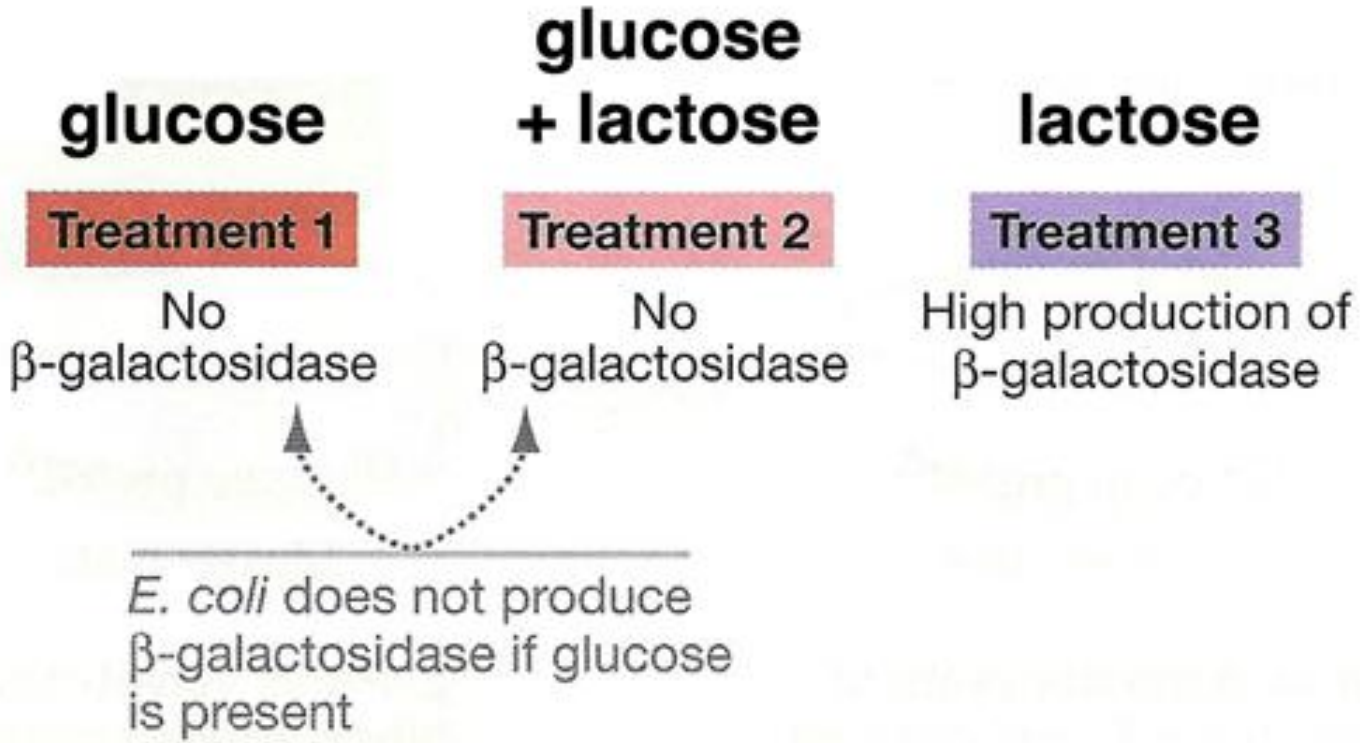




X-gal identifies cells with beta-galactosidase

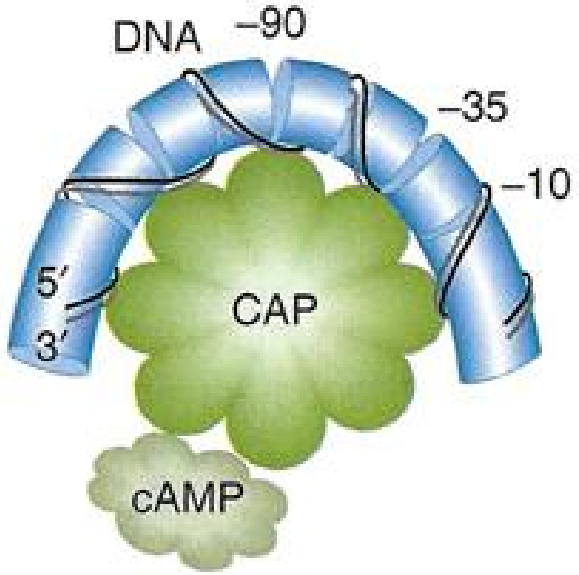


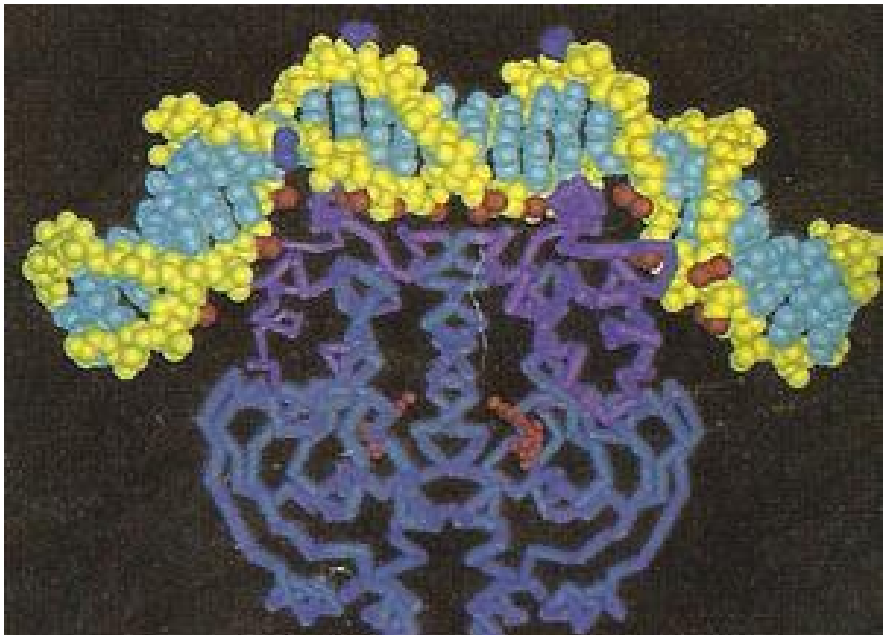
How does *E. coli* respond to lactose?



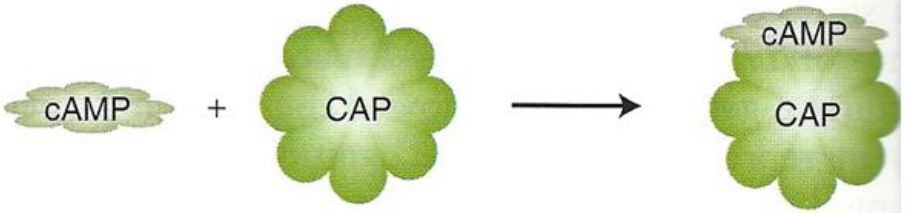
CONCLUSION: Glucose prevents expression of the gene for β-galactosidase. The presence of lactose without glucose stimulates expression of that gene.

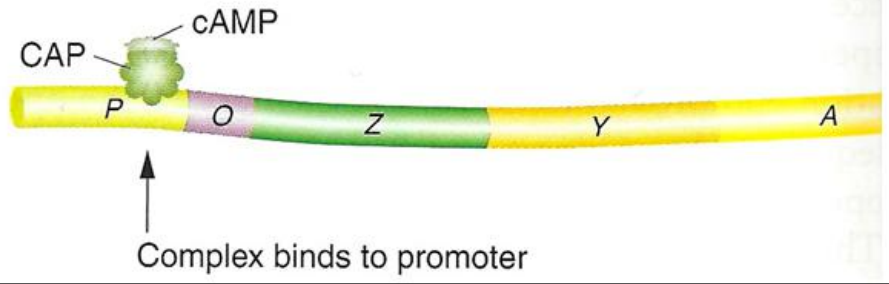
Activated CAP is required for promoter activation



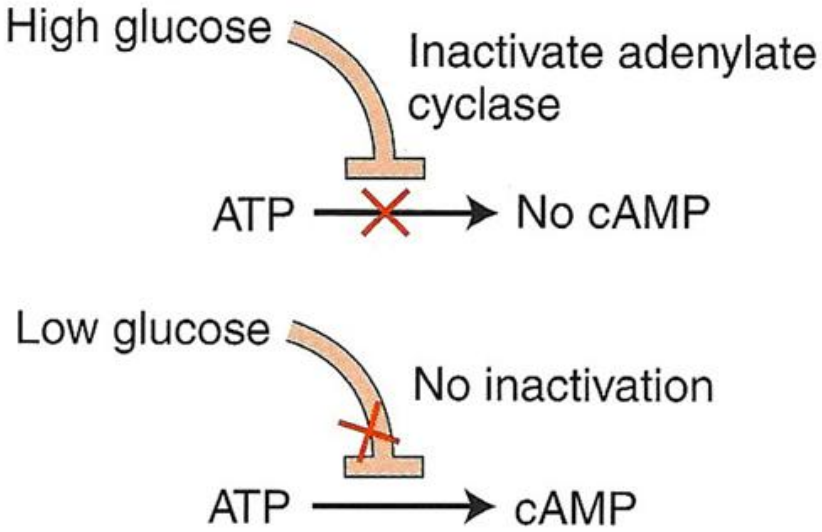


cAMP–CAP complex activates transcription

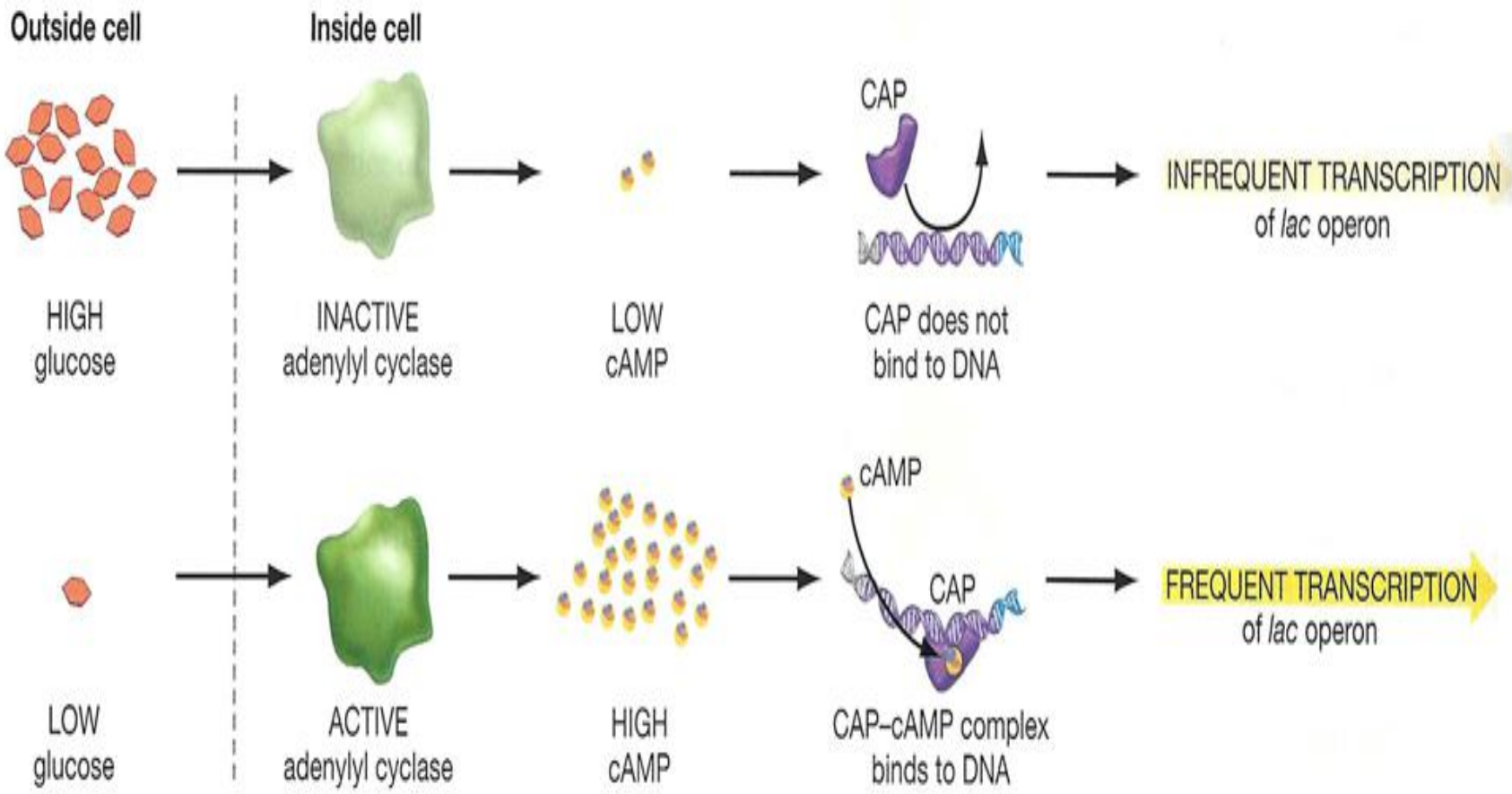




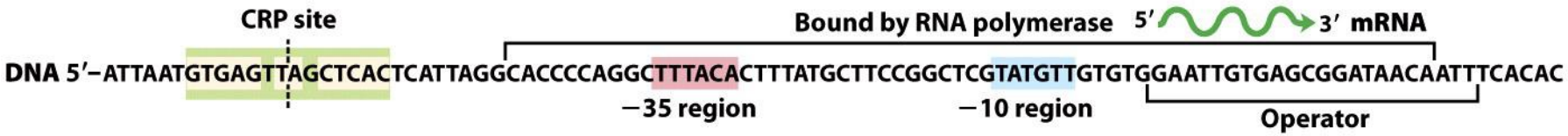
Glucose levels regulate cAMP levels

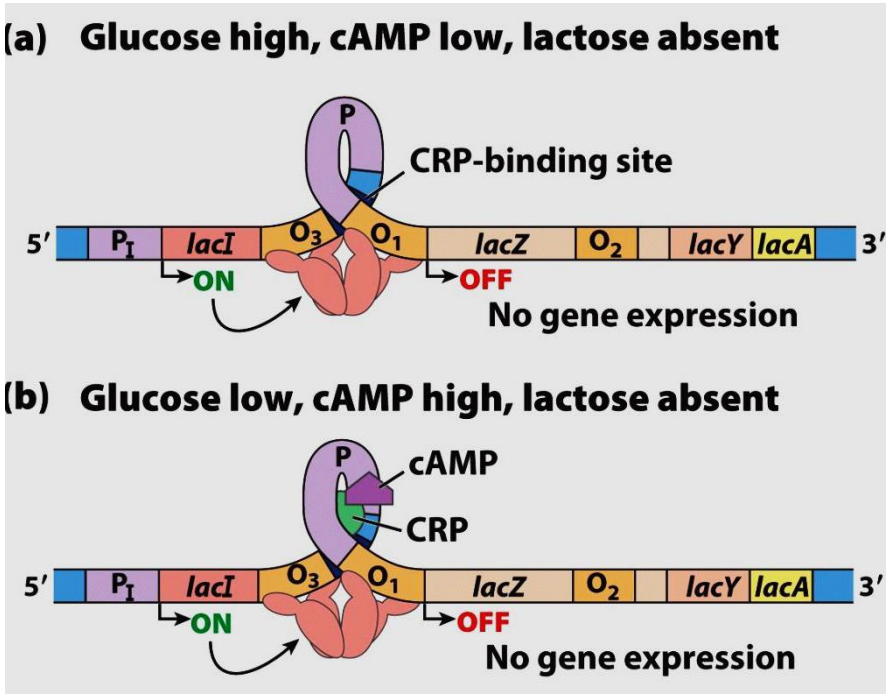


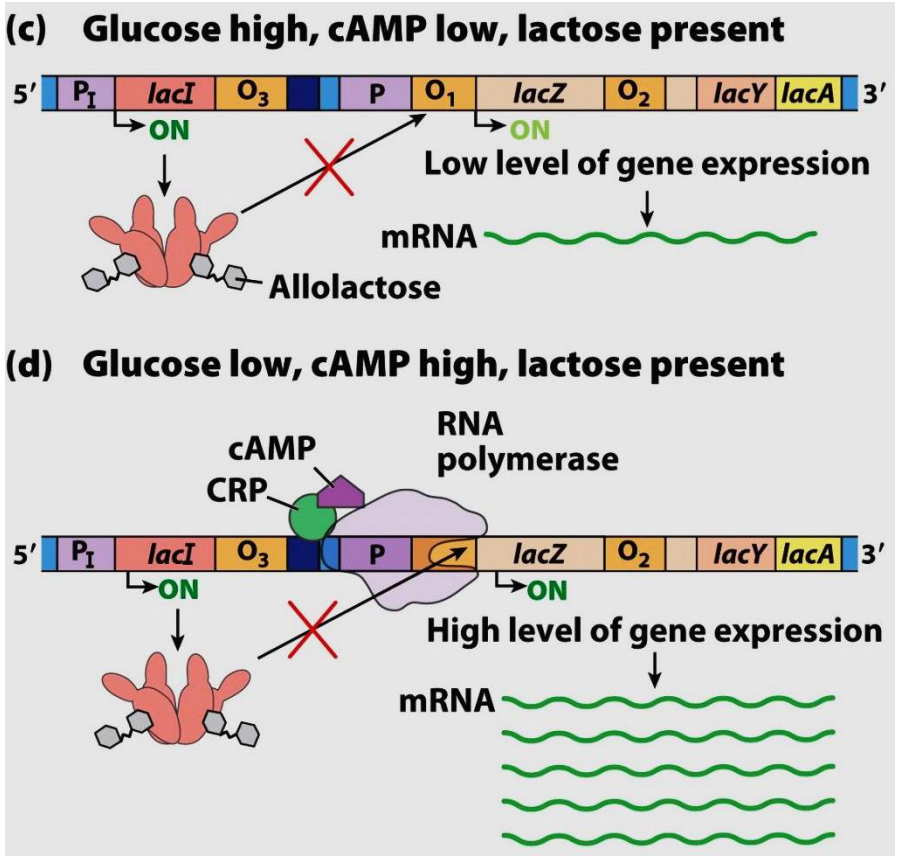
Glucose is priority

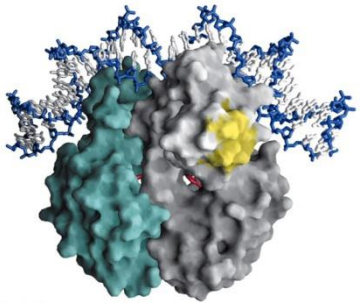


Combined effects of glucose and lactose on expression of the *lac* operon



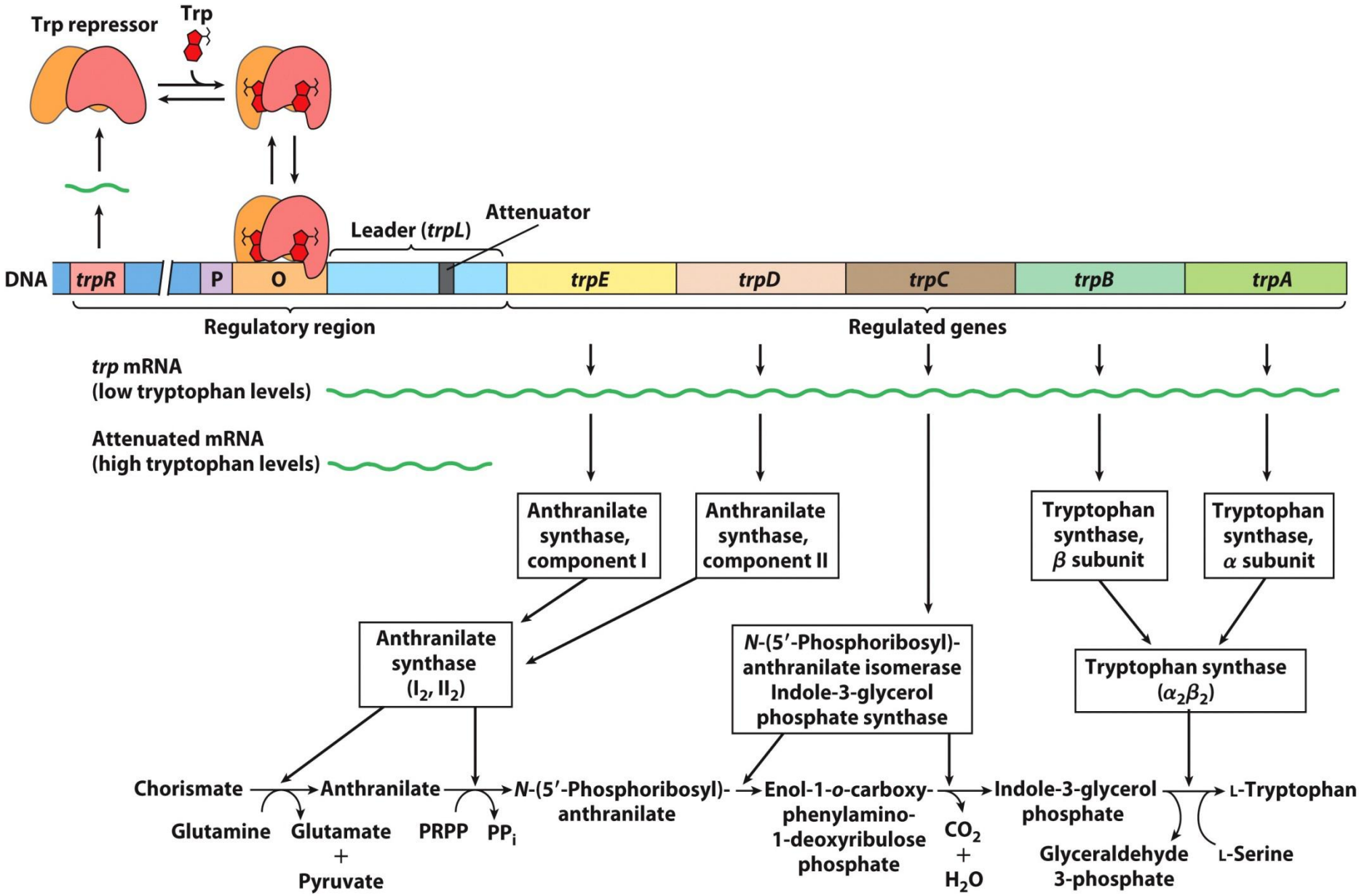






CRP: cAMP receptor protein

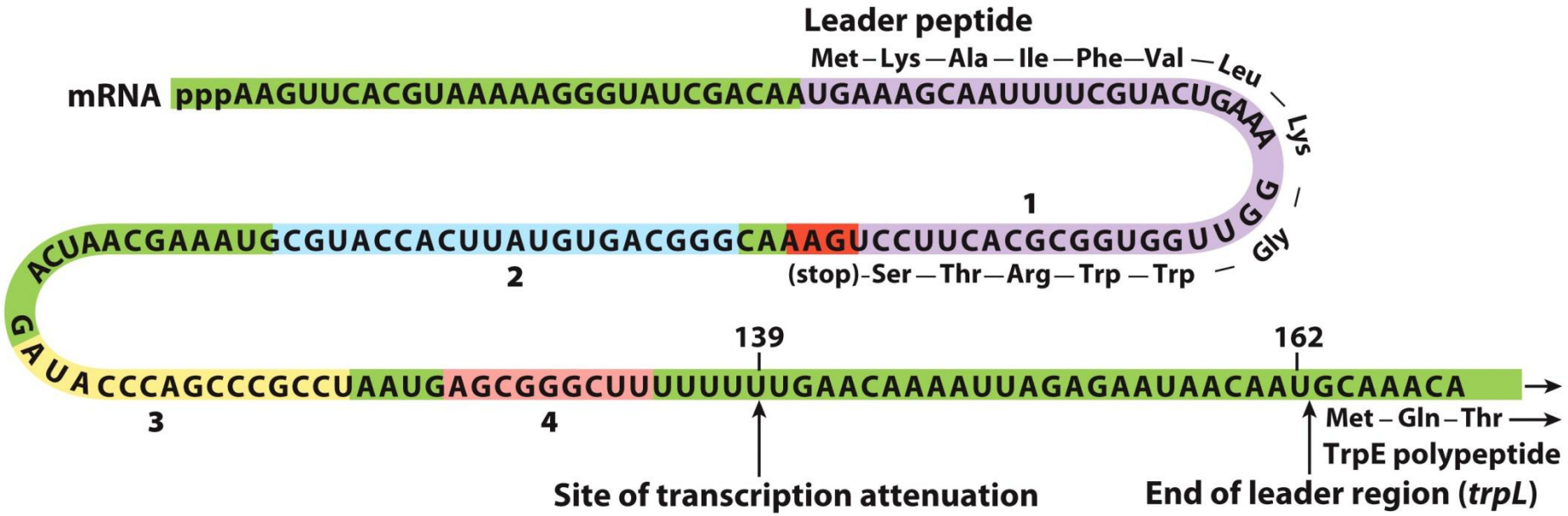
E. coli trp operon



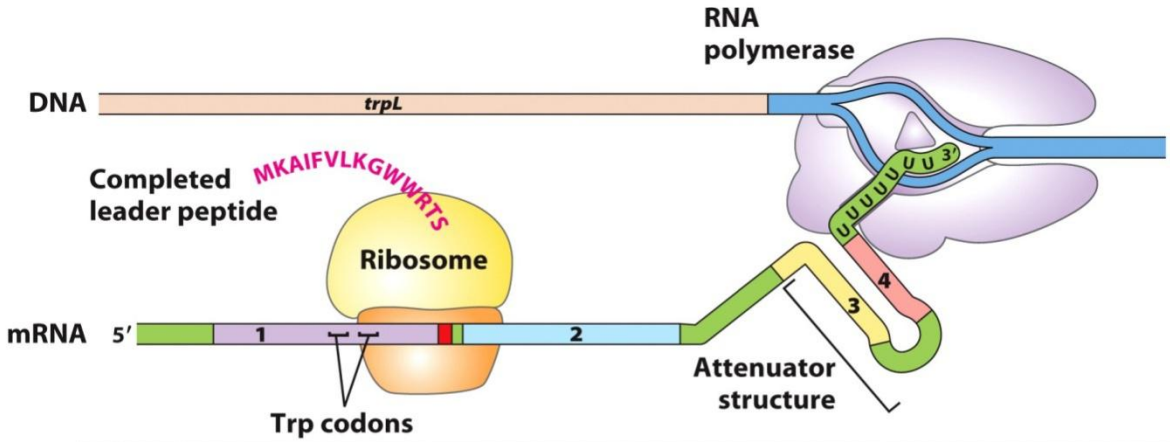
Transcriptional attenuation in the *trp* operon

Transcription attenuation: Transcription is initiated normally but is abruptly halted before operon genes are transcribed.

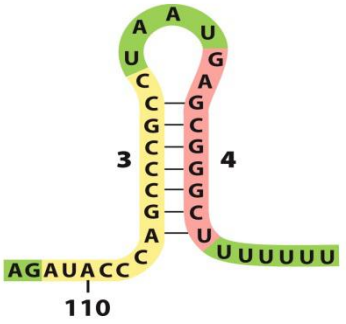
The *trp* mRNA leader



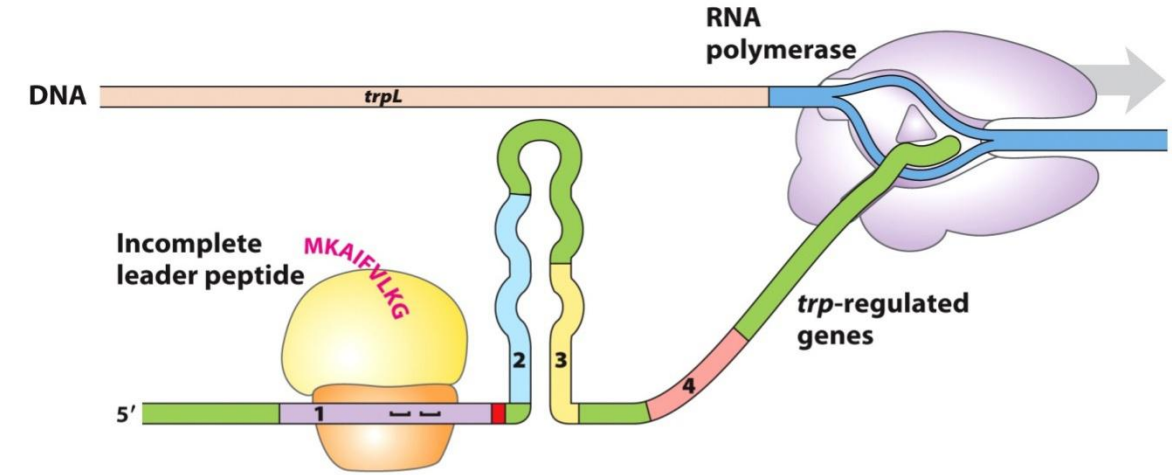
Transcriptional attenuation in the *trp* operon



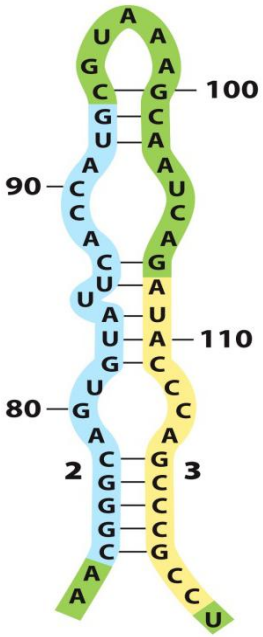
When tryptophan levels are high, the ribosome quickly translates sequence 1 (open reading frame encoding leader peptide) and blocks sequence 2 before sequence 3 is transcribed. Continued transcription leads to attenuation at the terminator-like attenuator structure formed by sequences 3 and 4.



3:4 Pair (attenuator)



When tryptophan levels are low, the ribosome pauses at the Trp codons in sequence 1. Formation of the paired structure between sequences 2 and 3 prevents attenuation, because sequence 3 is no longer available to form the attenuator structure with sequence 4. The 2:3 structure, unlike the 3:4 attenuator, does not prevent transcription.



2:3 Pair

Various amino acid biosynthetic operons regulated by attenuation

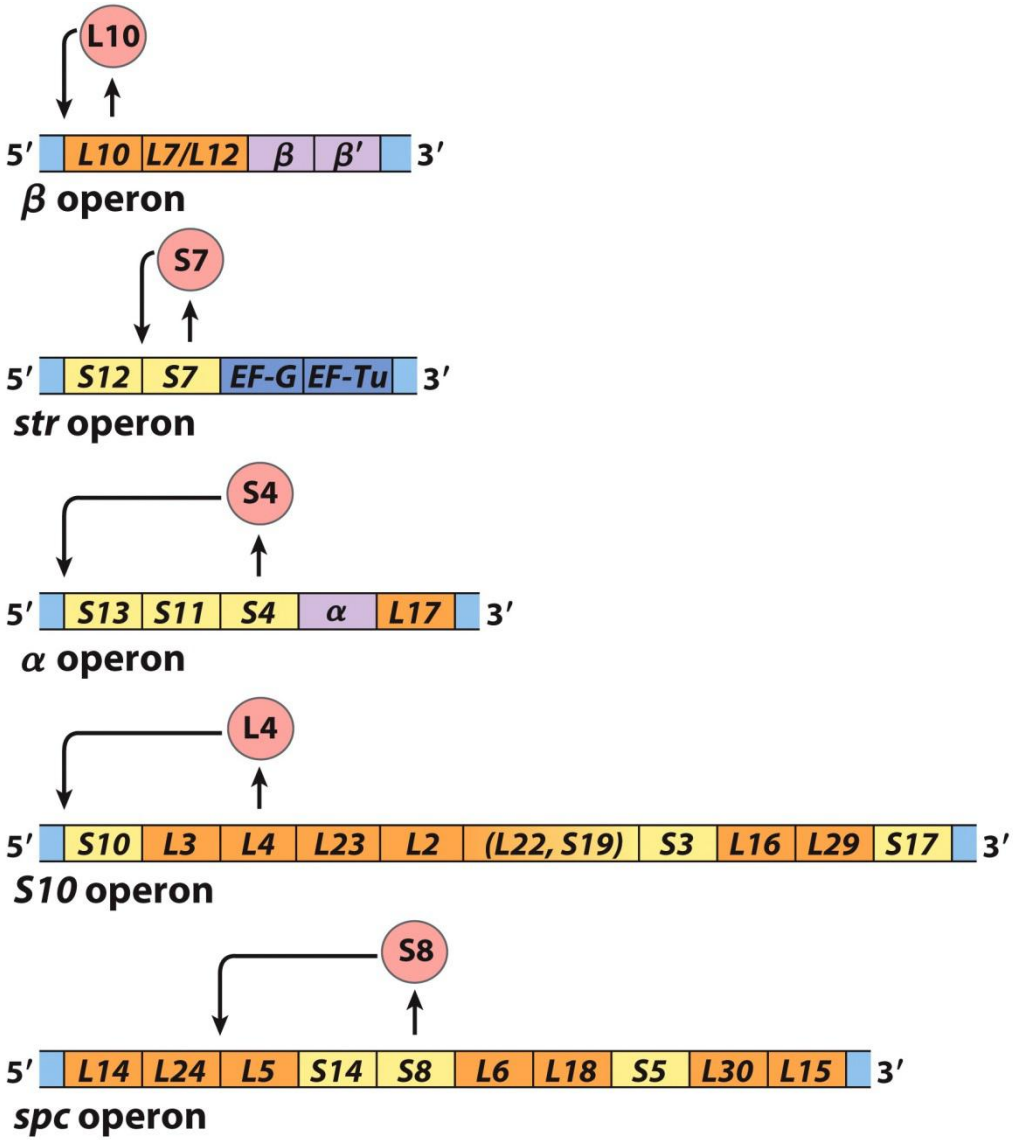
Operon	Amino acid Sequence
his	Met–Thr–Arg–Val–Glu–Phe–Lys–**His–His–His–His–His–His–His**–Pro–Asp
ilv	Met–Thr–Ala–**Leu–Leu**–Arg–**Val–Ile**–Ser–**Leu–Val–Val–Ile**–Ser–**Val–Val–Val–Ile–Ile–Ile**–Pro–Pro–Cys–Gly–Ala–Ala–Leu–Gly–Arg–Gly–Lys–Ala
leu	Met–Ser–His–Ile–Val–Arg–Phe–Thr–Gly–**Leu–Leu–Leu–Leu**–Asn–Ala–Phe–Ile–Val–Arg–Gly–Arg–Pro–Val–Gly–Gly–Ile–Gln–His
pheA	Met–Lys–His–Ile–Pro–**Phe–Phe–Phe**–Ala–**Phe–Phe–Phe**–Thr–**Phe**–Pro
thr	Met–Lys–Arg–**Ile**–Ser–**Thr–Thr**–Ile–**Thr–Thr–Thr**–Ile–**Thr**–Ile–**Thr–Thr**–Gln–Asn–Gly–Ala–Gly
trp	Met–Lys–Ala–Ile–Phe–Val–Leu–Lys–Gly–**Trp–Trp**–Arg–Thr–Ser

Translational feedback in some ribosomal protein operons

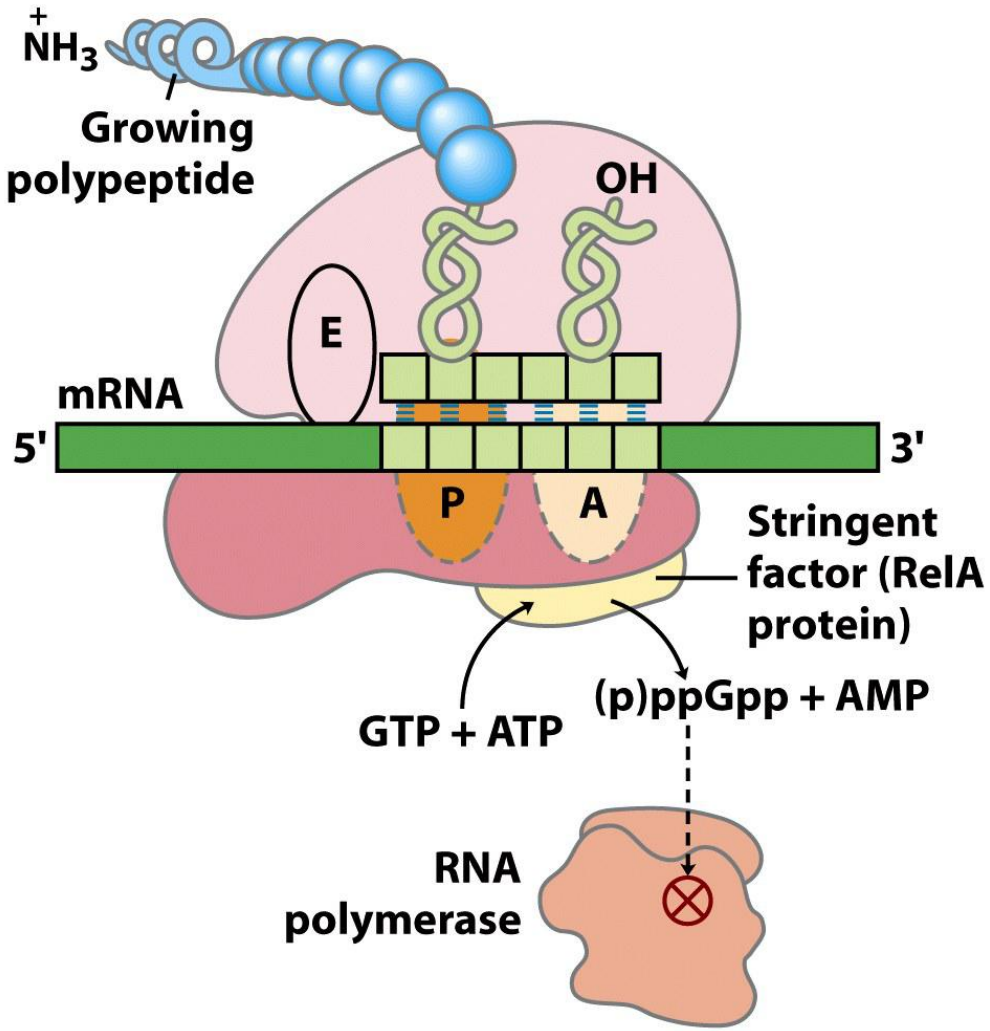
Synthesis of ribosomal proteins is coordinated with rRNA synthesis through translational repression

L1-L34: ribosomal proteins of the large subunit (50S=5S+23S +33 proteins)

S1-S21: ribosomal proteins of the small subunit (30S=16S +21 proteins)



Stringent response in *E. coli*



Amino acid starvation

↓

Binding of **stringent factor (RelA)** to the ribosome

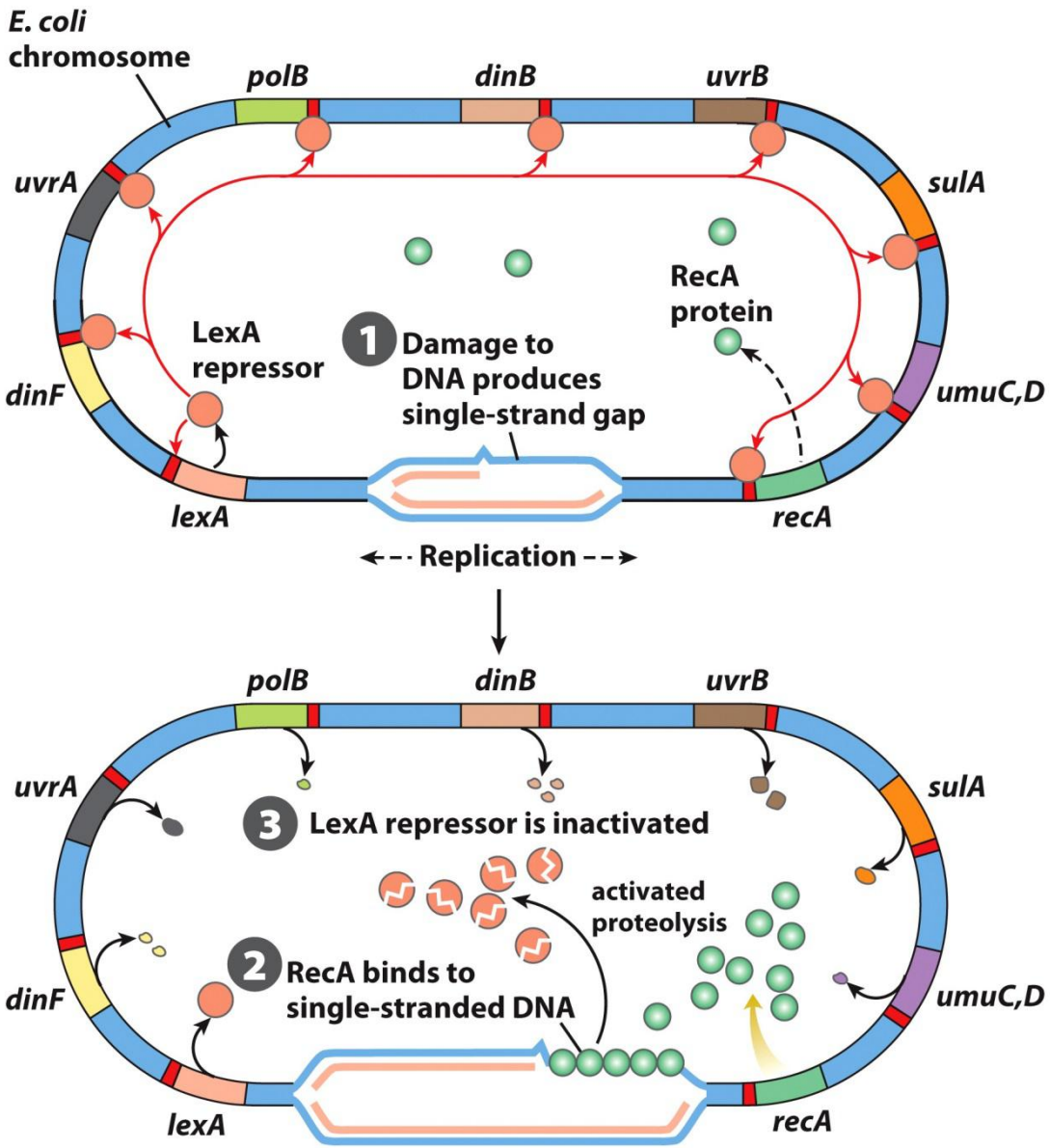
↓

Catalyzes formation of **guanosine tetraphosphate (ppGpp)**

↓

Inhibits RNA polymerase activity

SOS response induced gene expression in bacteria



SOS response:

Extensive DNA damage in bacterial chromosome triggers the induction of many distantly located genes.

Difference between regulation of transcription initiation in eukaryotes and prokaryotes

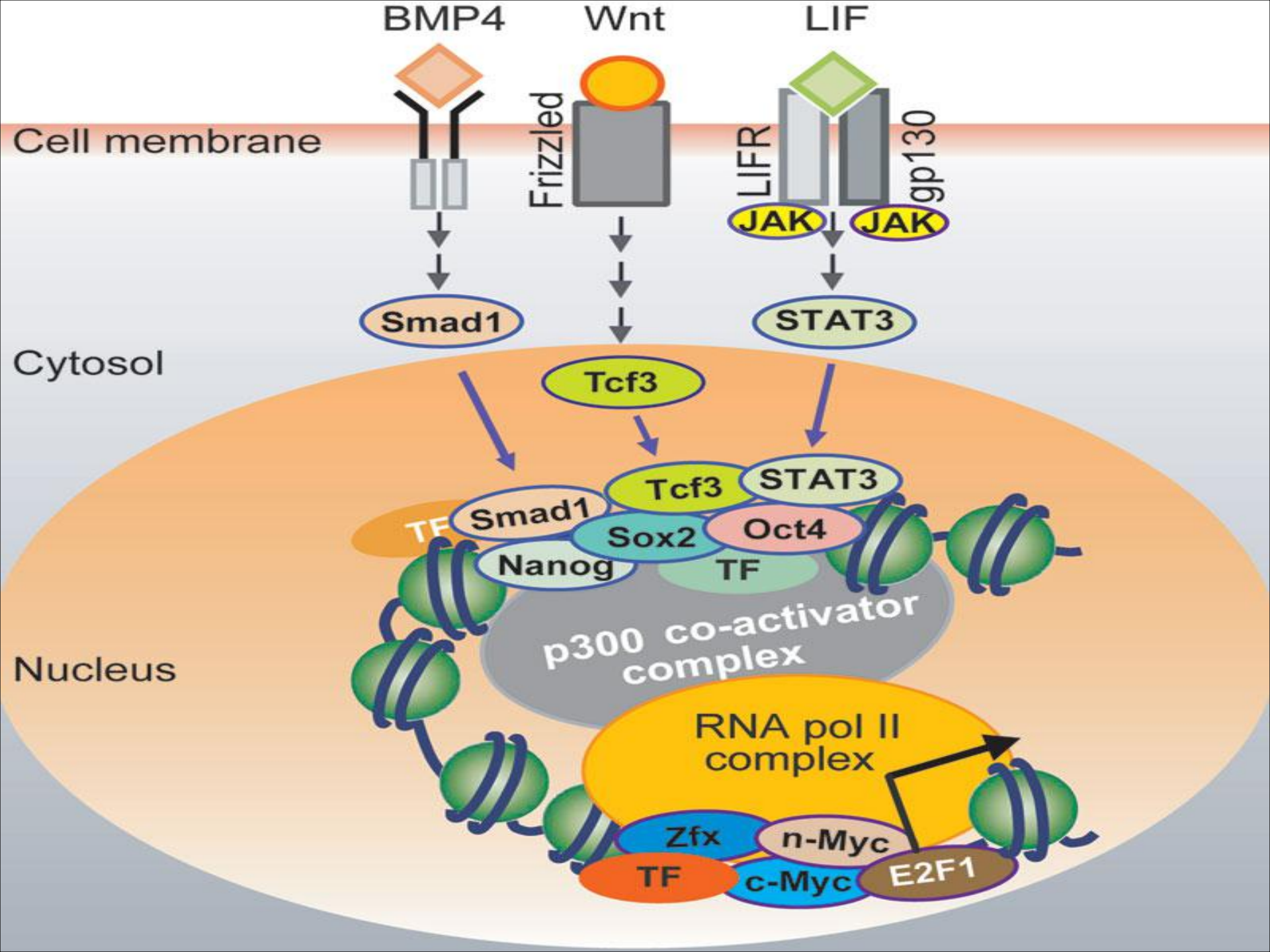
Bacteria:

The transcriptional is **nonrestrictive**, RNA polymerase can access to every promoter and initiate transcription at some level of efficiency in the absence of activators or repressors.

Eukaryotes:

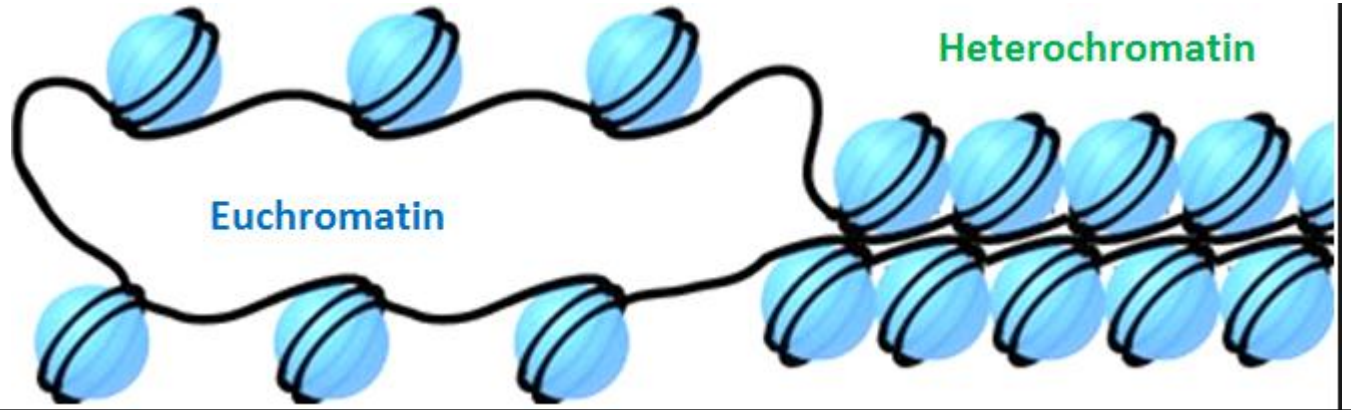
The transcriptional is **restrictive**. In the absence of transcription factors bound to the promoter, RNA polymerase ***cannot*** bind and initiate transcription (exceptions exist).

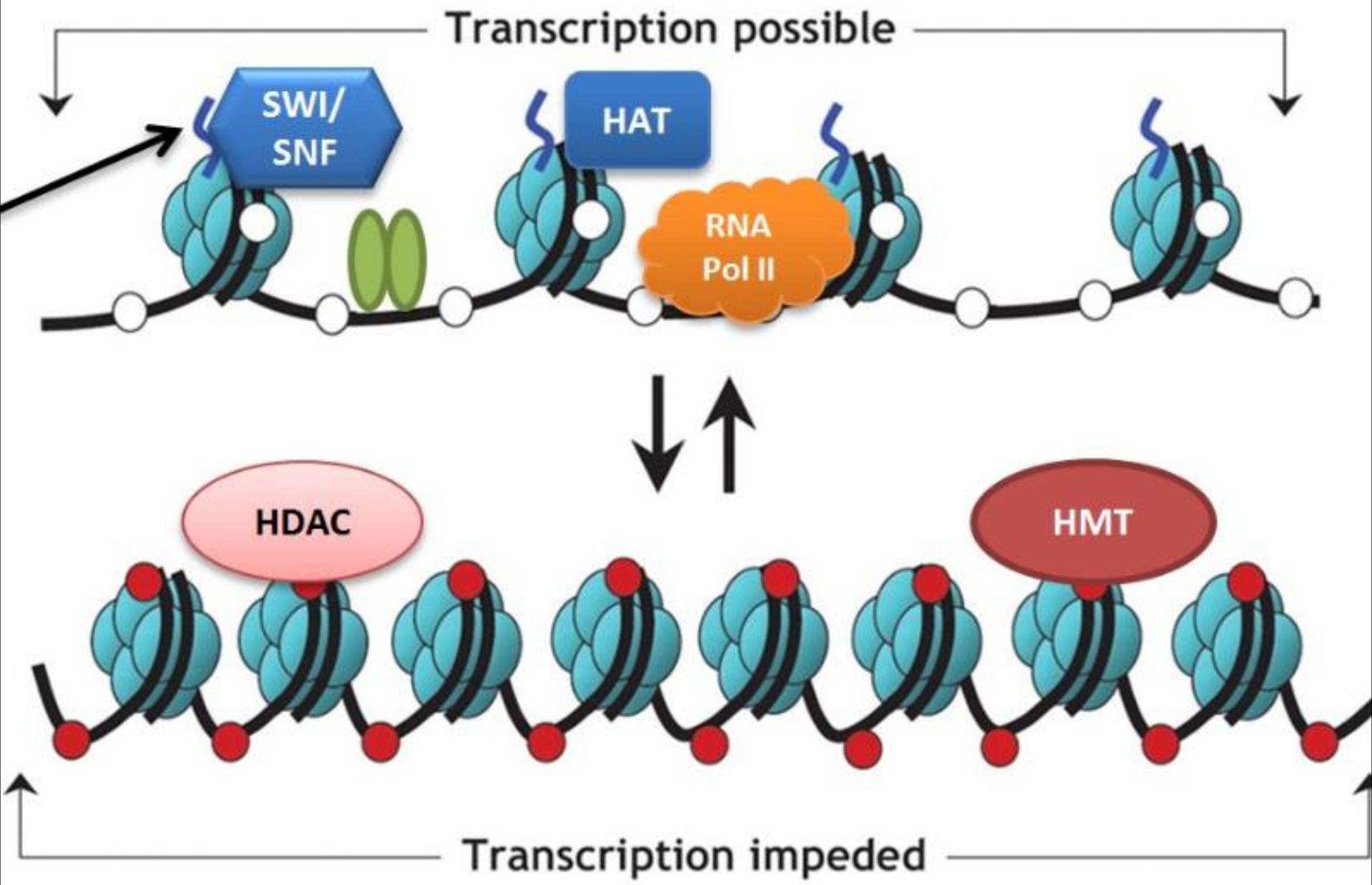
- Access to promoter DNA is restricted by the presence of chromatin.
- Protein activators of transcription initiation predominate in eukaryotes. More complicated, multimeric regulatory proteins are observed in eukaryotes.
- In eukaryotes, transcription and translation are separated in space and time.


BMP4
Wnt
LIF
Cell membrane
Frizzled
LIFR
gp130
JAK
JAK
Smad1
STAT3
Cytosol
Tcf3
Tcf3
STAT3
TF
Smad1
Sox2
Oct4
Nanog
TF
p300 co-activator complex
Nucleus
RNA pol II complex
Zfx
n-Myc
TF
c-Myc
E2F1

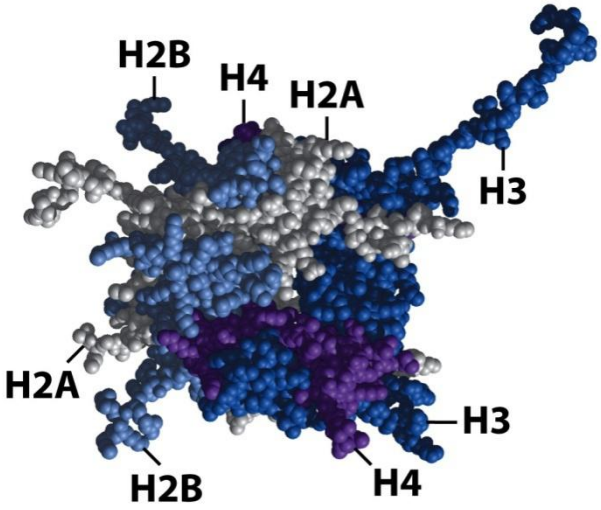
Chromosome structure and gene regulation

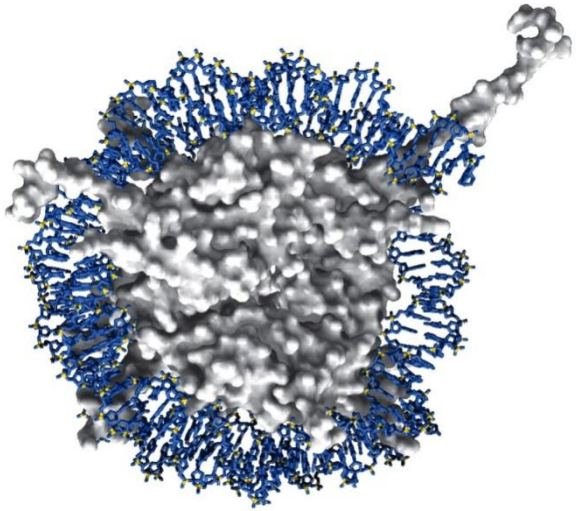
- **Heterochromatin**
  - Highly condensed DNA
  - Transcriptionally inactive
  - Associated with repeated sequences and structures such as the centromere
  - About 10% of total chromatin
- **Euchromatin**
  - Less condensed than heterochromatin
  - Some, but not all euchromatin, is transcriptionally active
  - About 90% of total chromatin

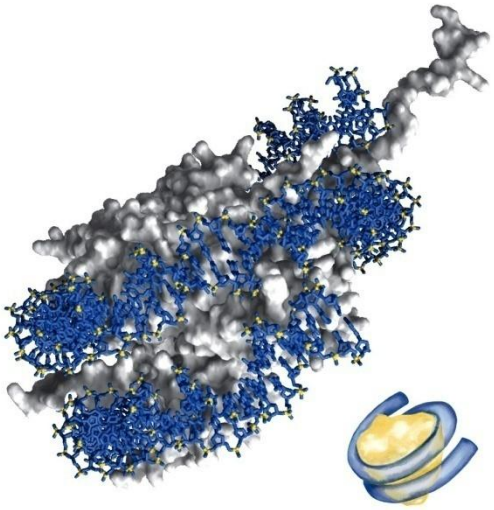


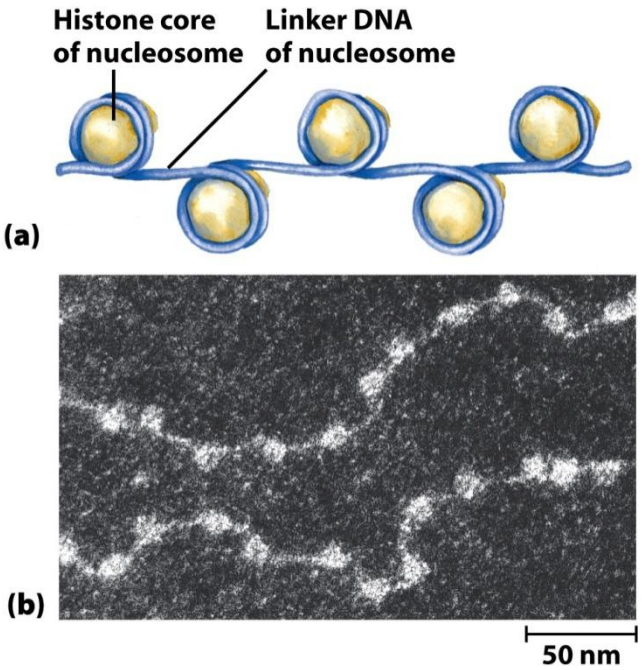

Transcription possible
SWI/
SNF
HAT
RNA
Pol II
HDAC
HMT
Transcription impeded

Histone and Nucleosome





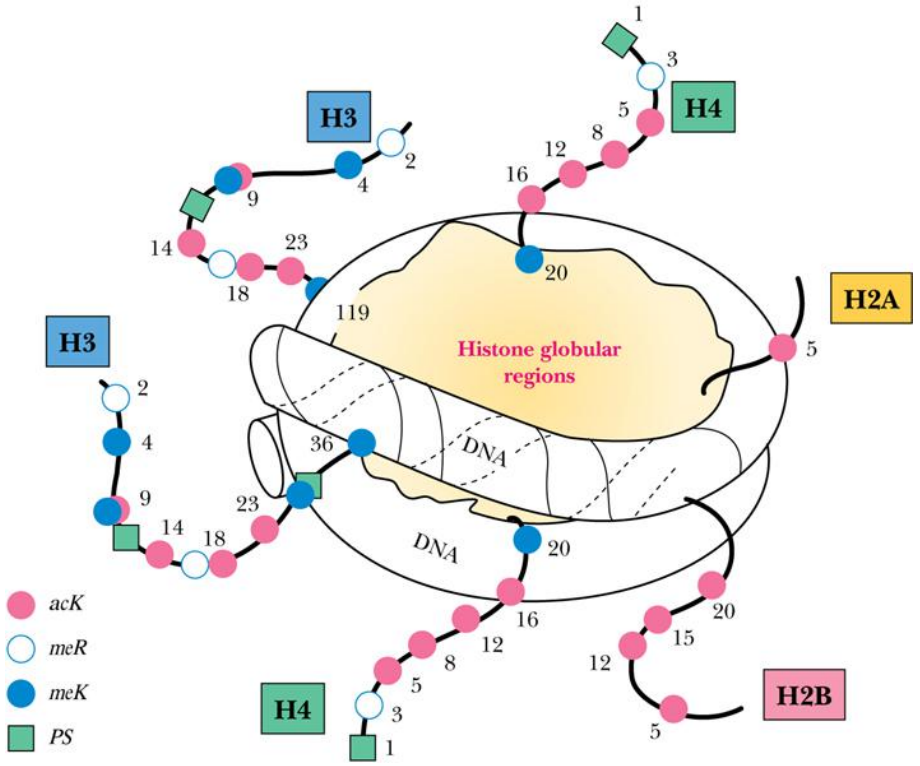




- The nucleosome not only serves to compact the genetic material but also provides information that affects nuclear functions including DNA replication, repair and transcription.

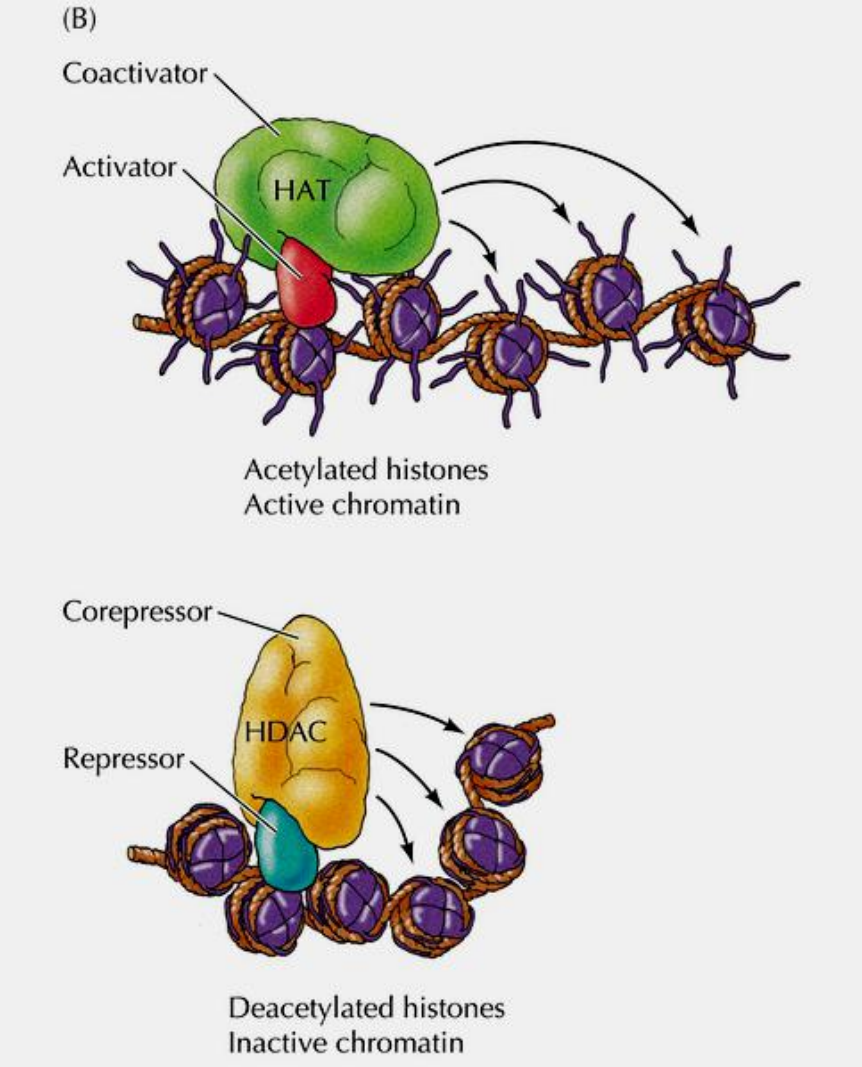
Dynamic Change of Chromatin Structure

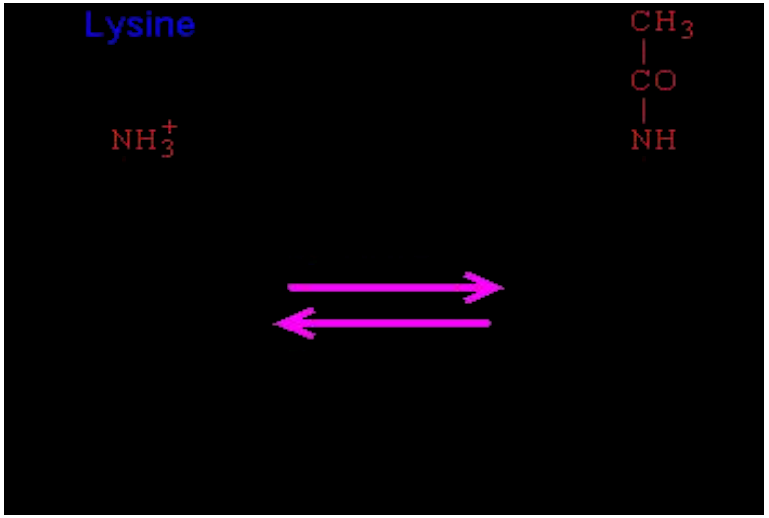
- Core histones are often post-translationally modified near their N-termini.
  - Methylation of lysines
  - Ser/Thr phosphorylation
  - Acetylation
  - Ubiquitination
  - Sumolyation
  - …



- Modifications on histones cause structural changes in chromatin

Chromatin Remodeling



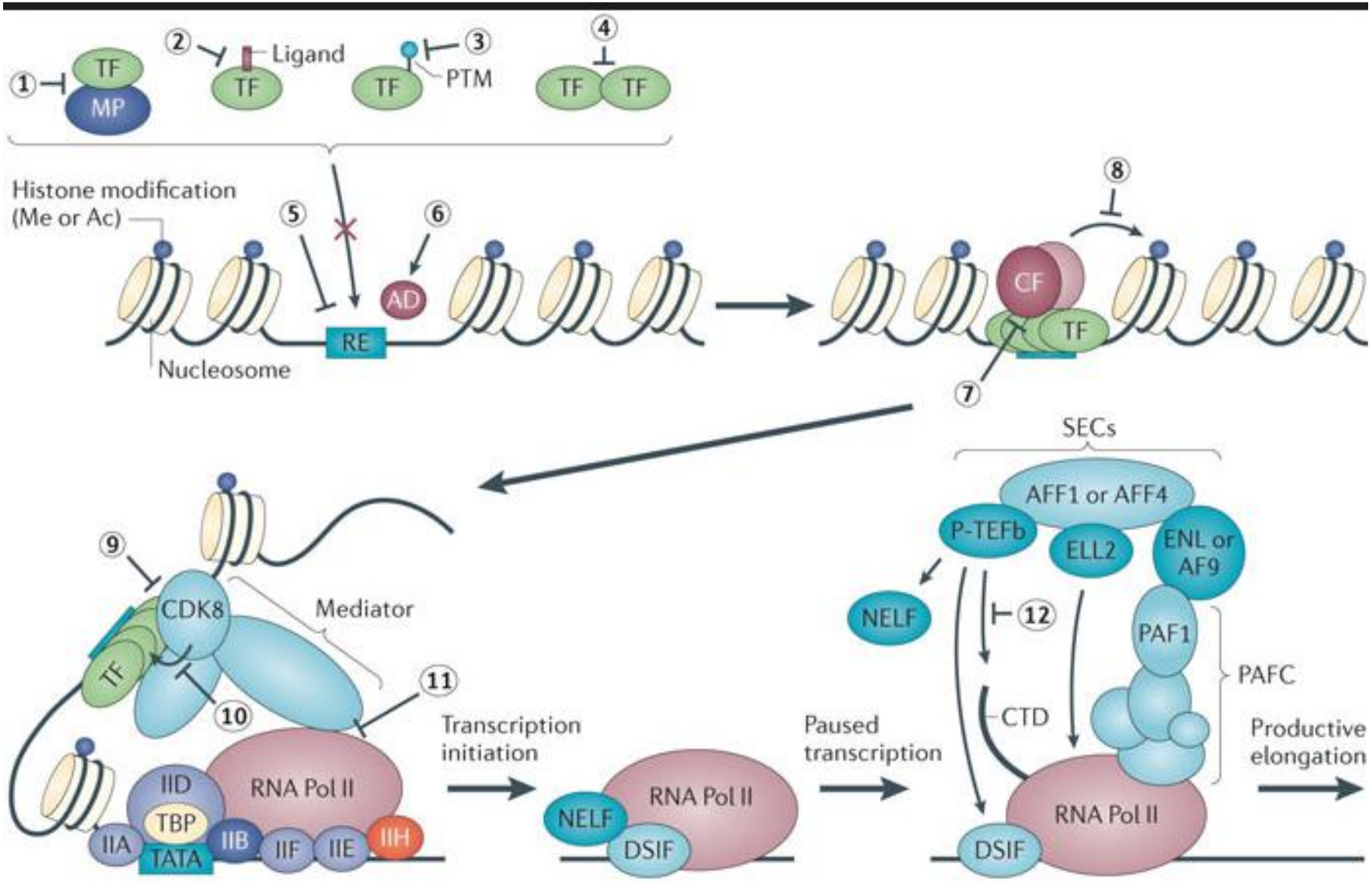


- Acetylation of histone reduces the affinity of nucleosomes for DNA and predisposes the DNA for transcription.
- De-acetylation "silences" the chromatin by reducing the frequency of transcription initiation.

Histone acetylation is catalyzed by **histone acetyltransferases (HATs)** ,

Histone deacetylation is catalyzed by **histone deacetylases** (**HDs** or **HDACs**).

Chromatin and transcription



Epigenetics

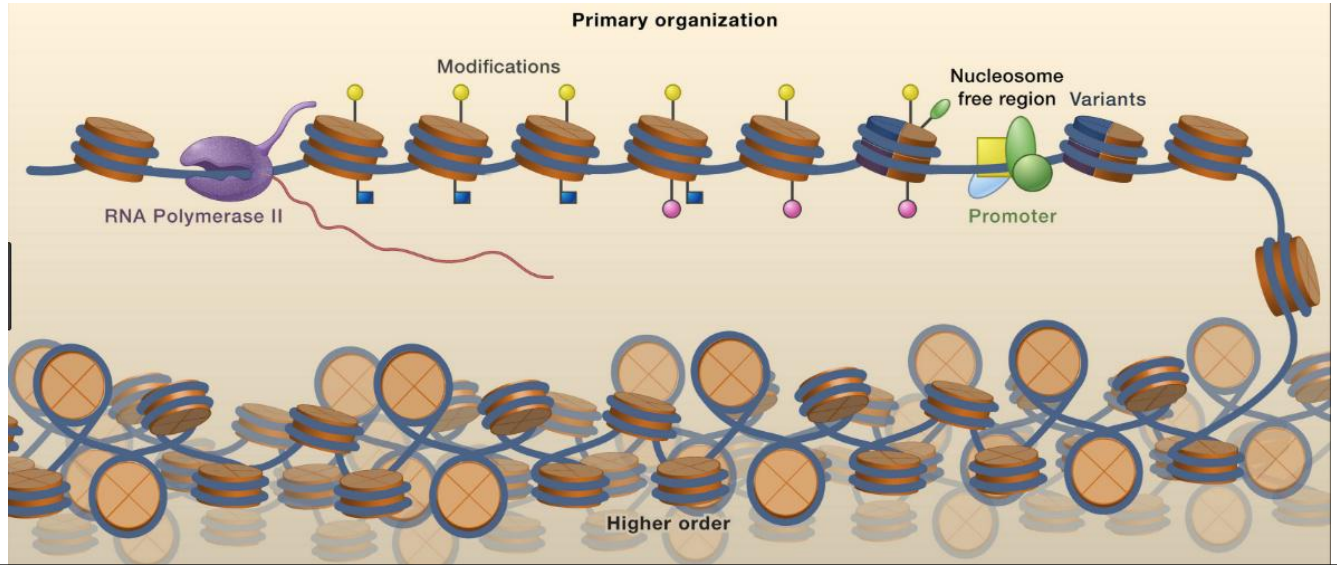
Epigenetics can be defined as any inheritable influence on gene activity that does not involve a change in DNA sequence.

Common mechanisms may include but not limited to:

-DNA methylation

-Histone modification/histone variants

-Regulatory non-coding RNAs



Regulation of transcription initiation in eukaryotes

Five types of proteins are required for successful binding of RNA polymerase II (RNA Pol II) holoenzyme at promoter:

(1) **Transcription activators**: bind to enhancers or upstream activator sequences (UASs) to facilitate transcription;

(2) **Architectural regulators**: facilitate DNA looping;

(3) **Chromatin modification and remodeling proteins**: modify histone to mediate chromatin structure;

(4) **Coactivators**: does not bind to the DNA, but are required for essential communication between the transcription activators and the complex of Pol II and the general transcription factors, such as **mediator**;

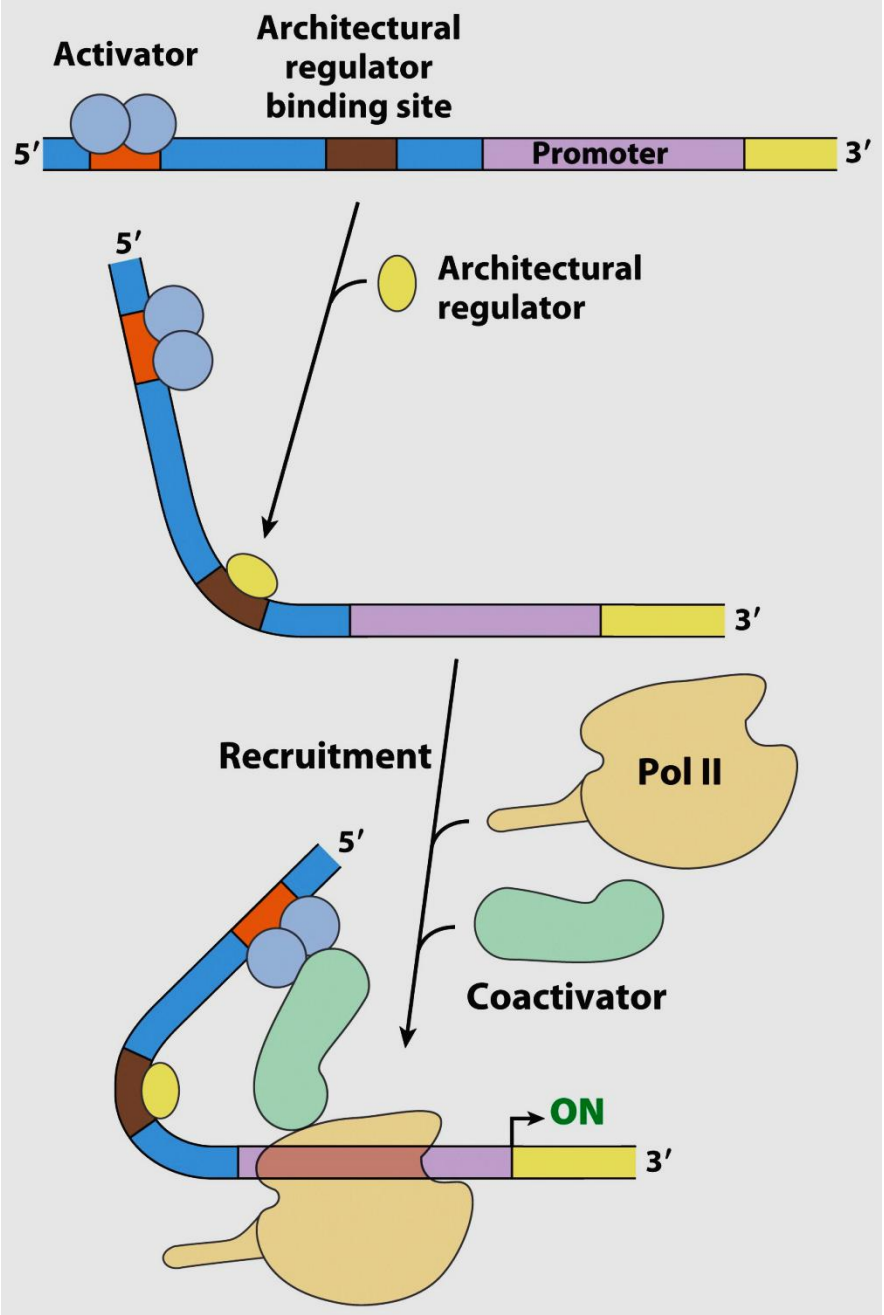
(5) **Basal transcription factors**: required at every Pol II promoter.

Repressors: repressor proteins can interfere with communication between the RNA polymerase and the transcription activators to repress transcription.

Architectural regulators in the initiation of transcription in eukaryotes

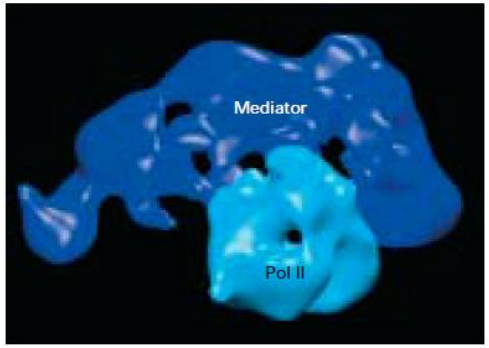
Architectural regulators facilitate the interaction between activators and RNA polymerase by looping out the DNA in between.

High mobility group (HMG) proteins

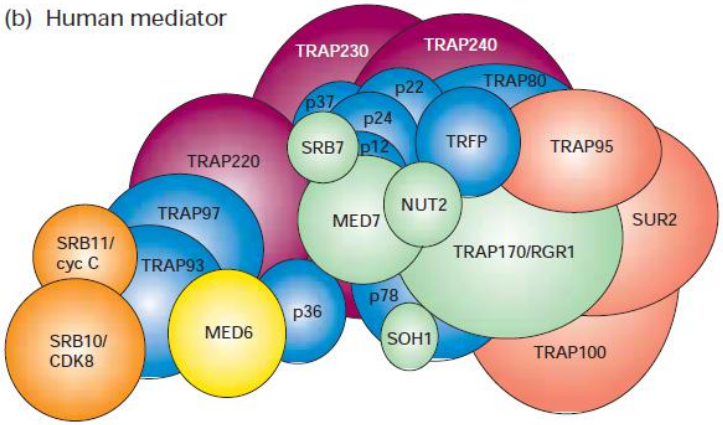


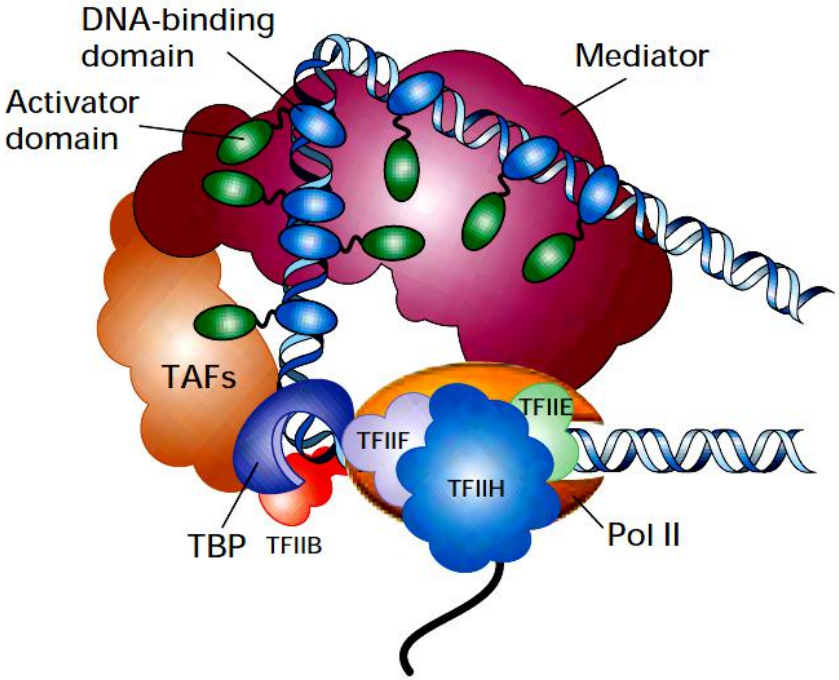
The mediator complex forms a molecular bridge between transcription activators and Pol II complex

(a) Yeast mediator–Pol II complex

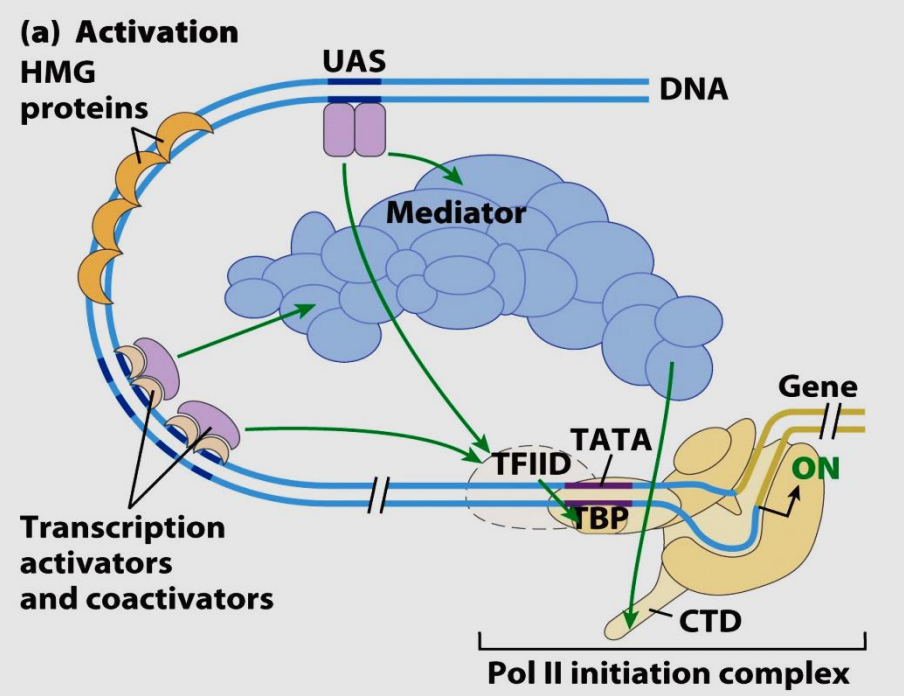


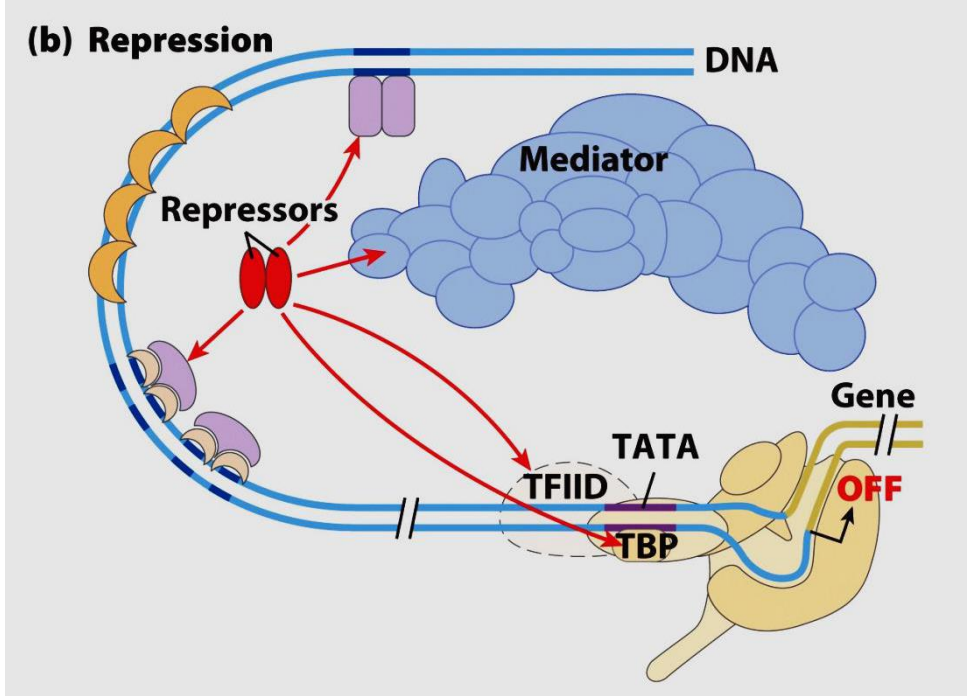
(b) Human mediator





Eukaryotic promoters and regulatory proteins

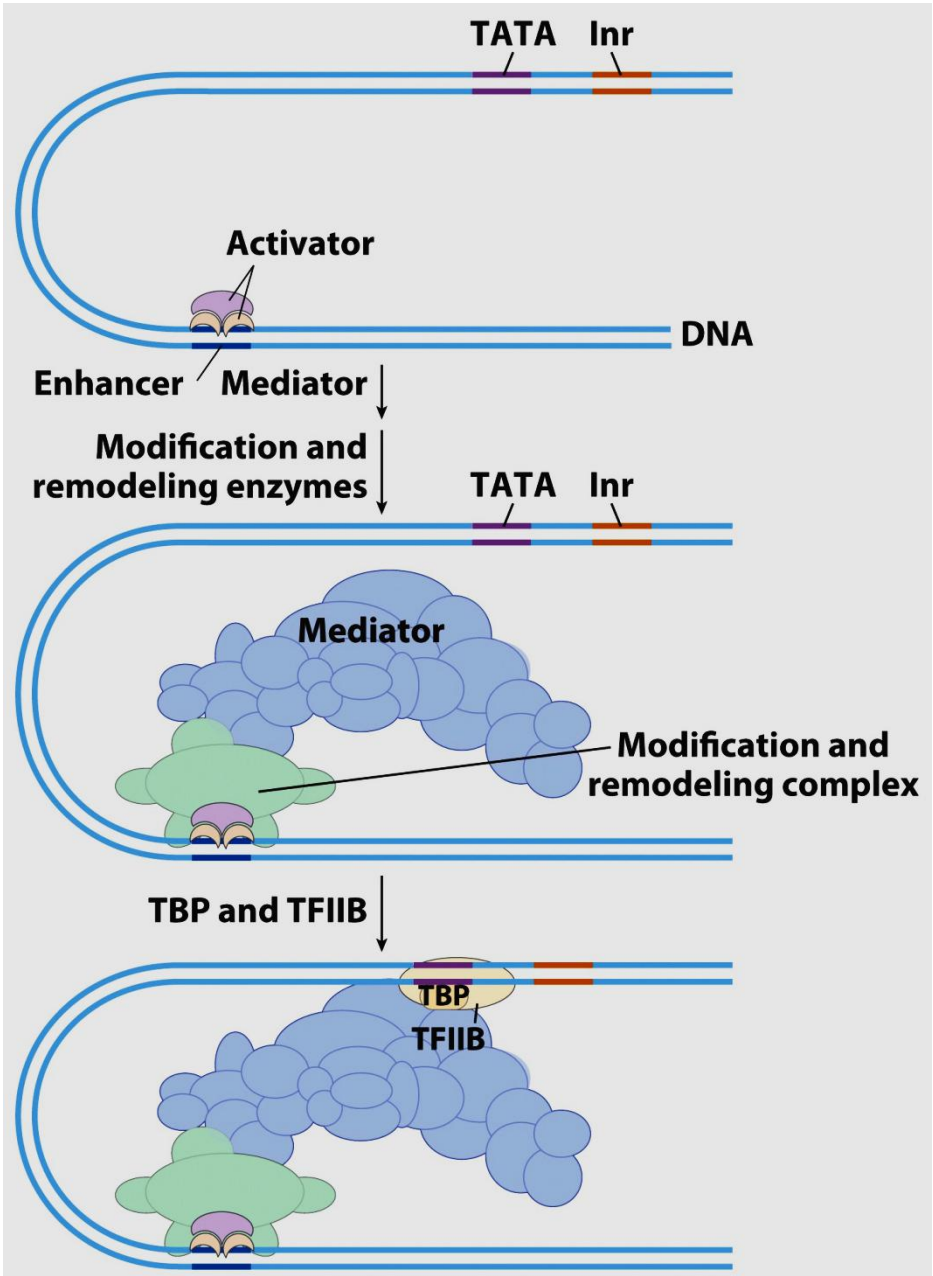


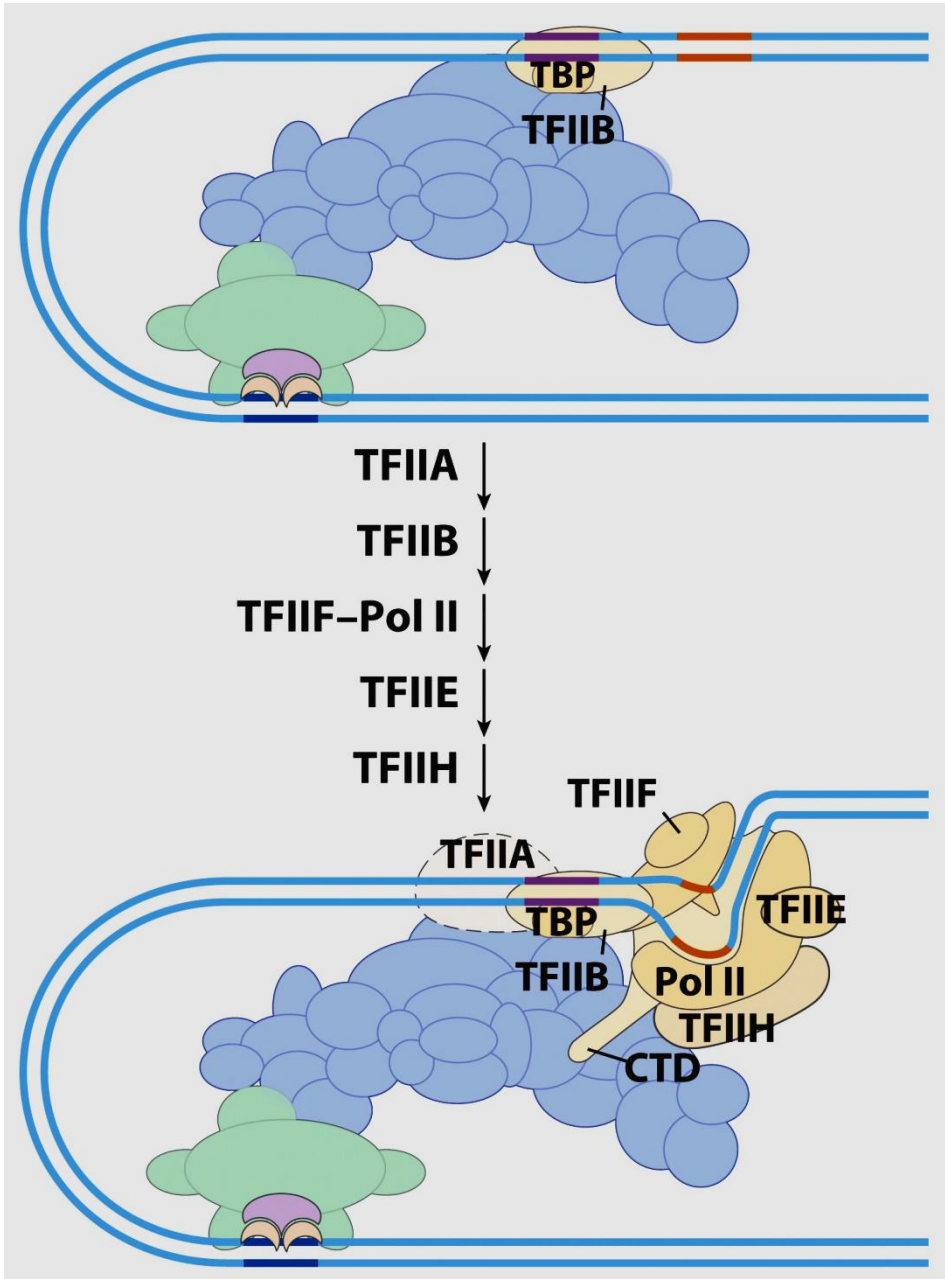


RNA polymerase II and its associated basal (general) transcription factors form a pre-initiation complex at the TATA box and Inr site of the promoters, facilitated by transcription activators (mediator, TFIID, or both).

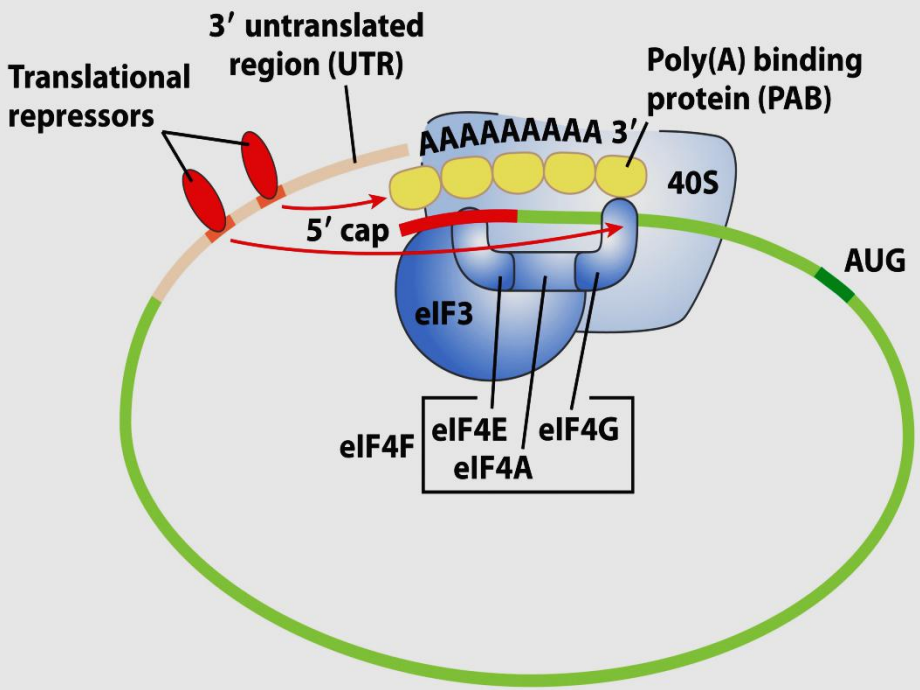
Eukaryotic transcriptional **repressors** function by a wide range of mechanisms. Some bind directly to DNA, displacing a protein complex required for activation; others interact with various parts of the transcription or activation complexes to prevent activation.

Schematic of transcriptional activation





Translational regulation of eukaryotic mRNA



Regulation of translation is more common in eukaryotes than prokaryotes

Four main mechanisms of translational regulation in eukaryotes:

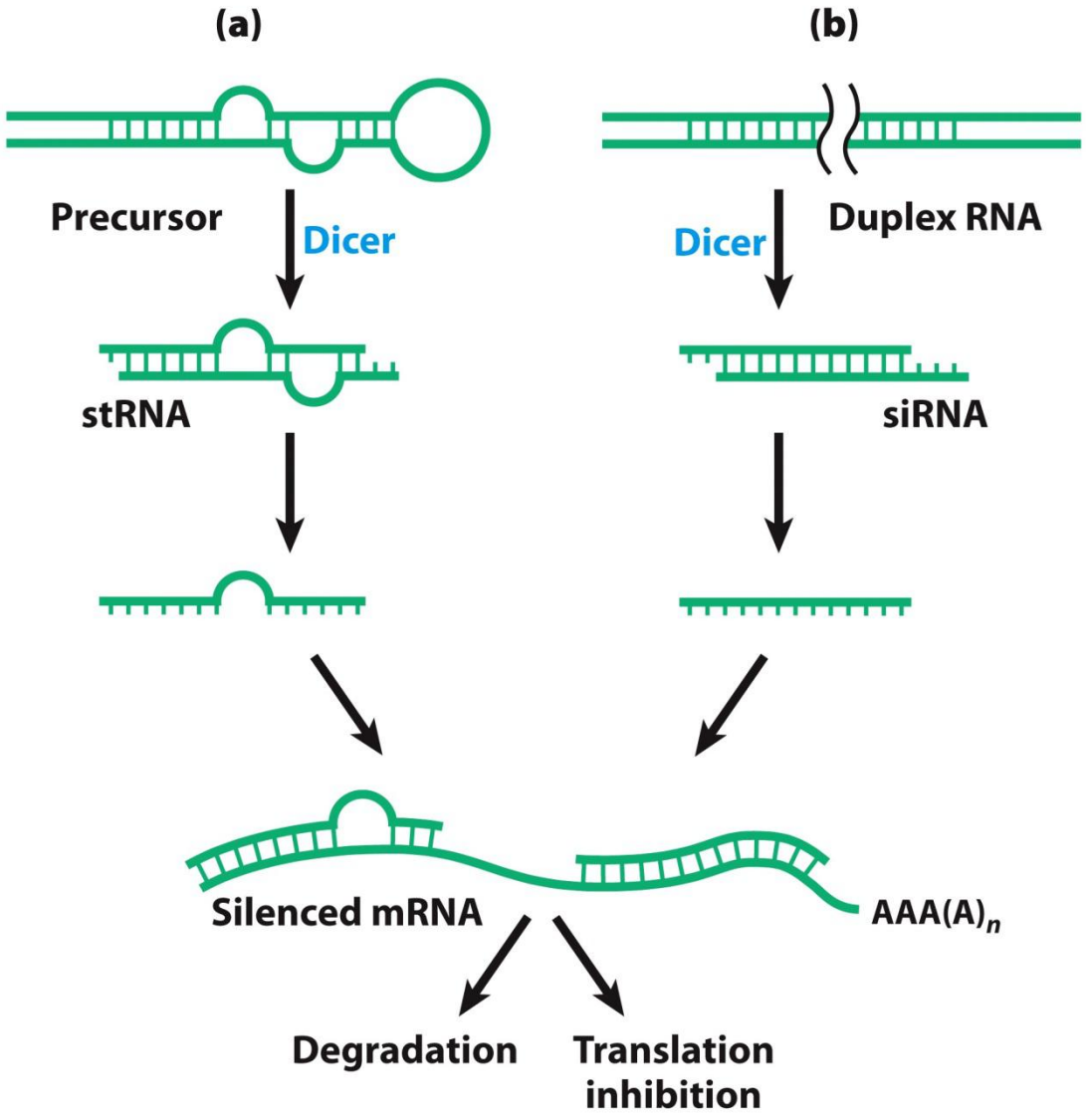
1) Phosphorylation of initiation factors: The phosphorylated forms are often less active and cause a general depression of translation in the cell.

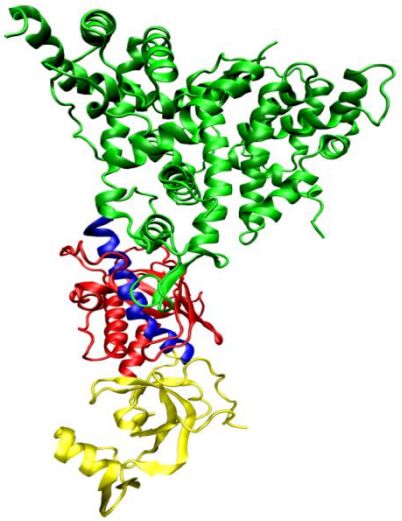
2) Binding directly to mRNA: Translational repressors bind directly to mRNA, frequently at specific sites in the 3’ untranslated region (3’-UTR).

3) Some binding proteins disrupt the interaction between eIF4E and eIF4G.

4) RNA-mediated gene expression regulation: siRNAs, and microRNAs

Gene silencing by RNA interference





Dicer cleaves double-stranded RNA (dsRNA) and pre-miRNA into short double-stranded miRNA about 20-25 nucleotides long, usually with a two-base overhang on the 3' end.

Artificially synthesized siRNA/shRNA can be used to shut down virtually any gene.

miRNAs are produced from either their own genes or from introns