

# **Cellular and Molecular Biology I**

**BIO-205-7**

**Camille Goemans**

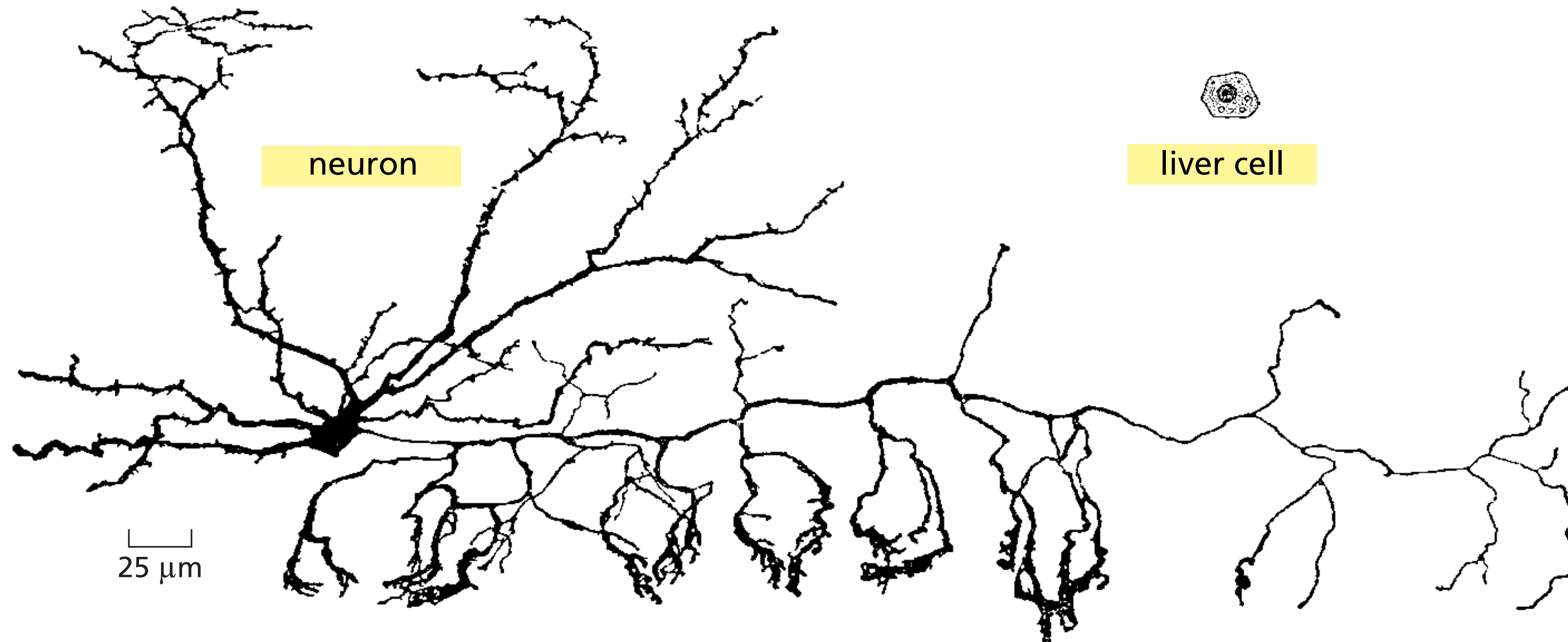
# IV. Control of gene expression

## 1. Transcriptional control

- a. Transcription regulators (activators, repressors)
- b. Understanding different control systems
- c. Combinatorial gene control and cell types

# Different cell types contain the same DNA

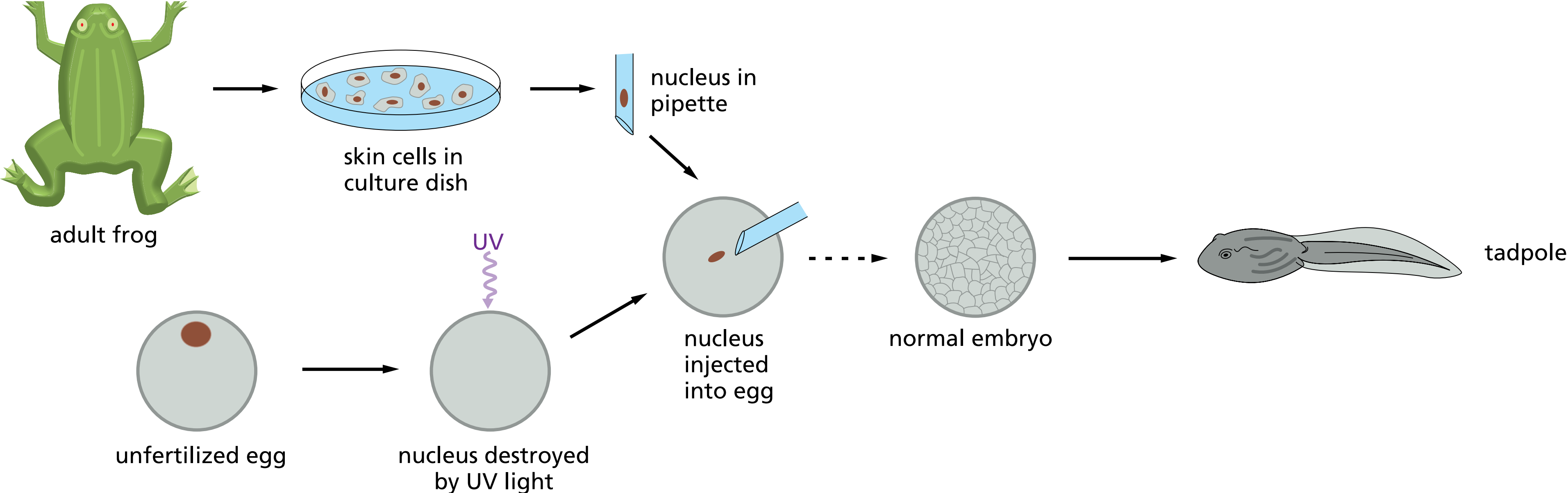
Yet, they are **different** and perform **different functions**



They express **different sets** of mRNA and proteins

# Different cell types contain the same DNA

How do we know this?

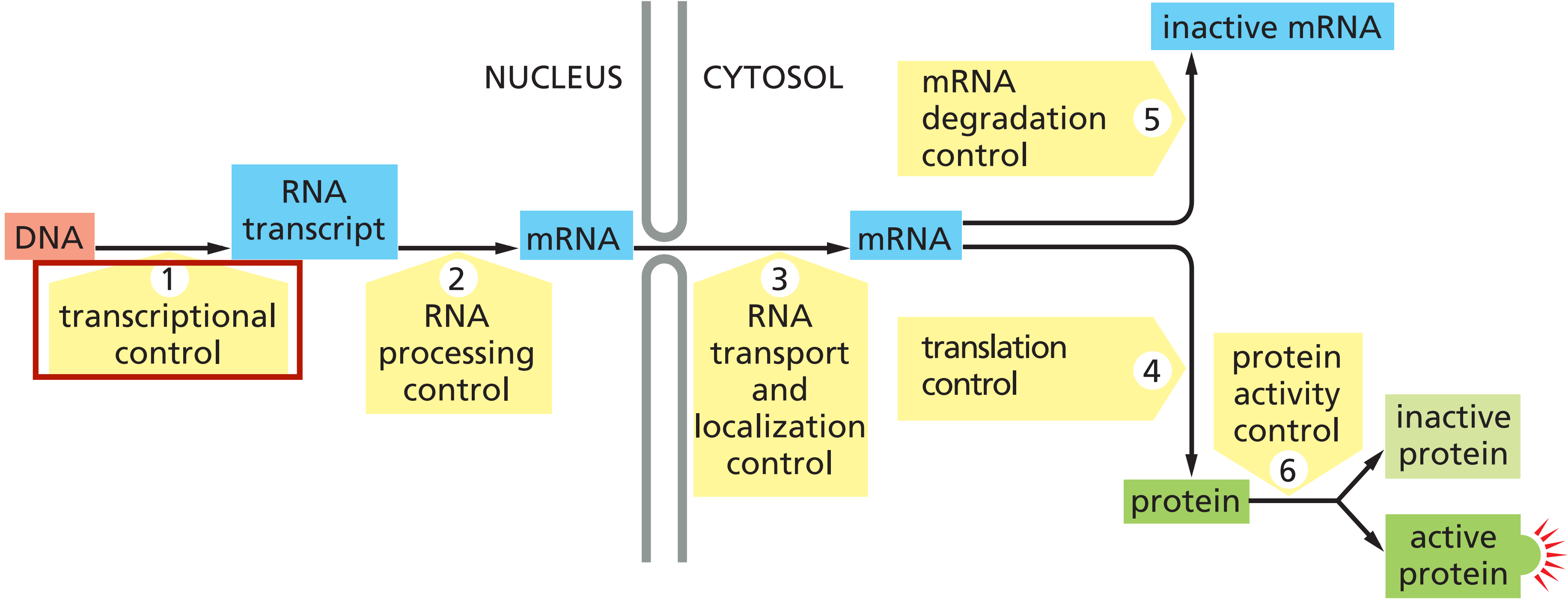


# Different cell types contain the same DNA

What are the **differences** between a cell type and another?

- Many processes are **common** (e.g. chromosomes, RNA and DNA polymerases, ribosomes, etc.)
- Some RNAs and proteins are abundant in **specialized cells**
- At any time, cells express **~30-60% of their ~ 30 000 genes**

# Levels of gene expression regulation



# IV. Control of gene expression

## 1. Transcriptional control

### **a. Transcription regulators (activators, repressors)**

b. Understanding different control systems

c. Combinatorial gene control and cell types

# Definitions

## Transcription Factors (TFs)

- Proteins that bind directly to specific DNA sequences, usually in promoter or enhancer/silencer regions, to control the initiation of transcription by RNA polymerase
- They can activate or repress transcription

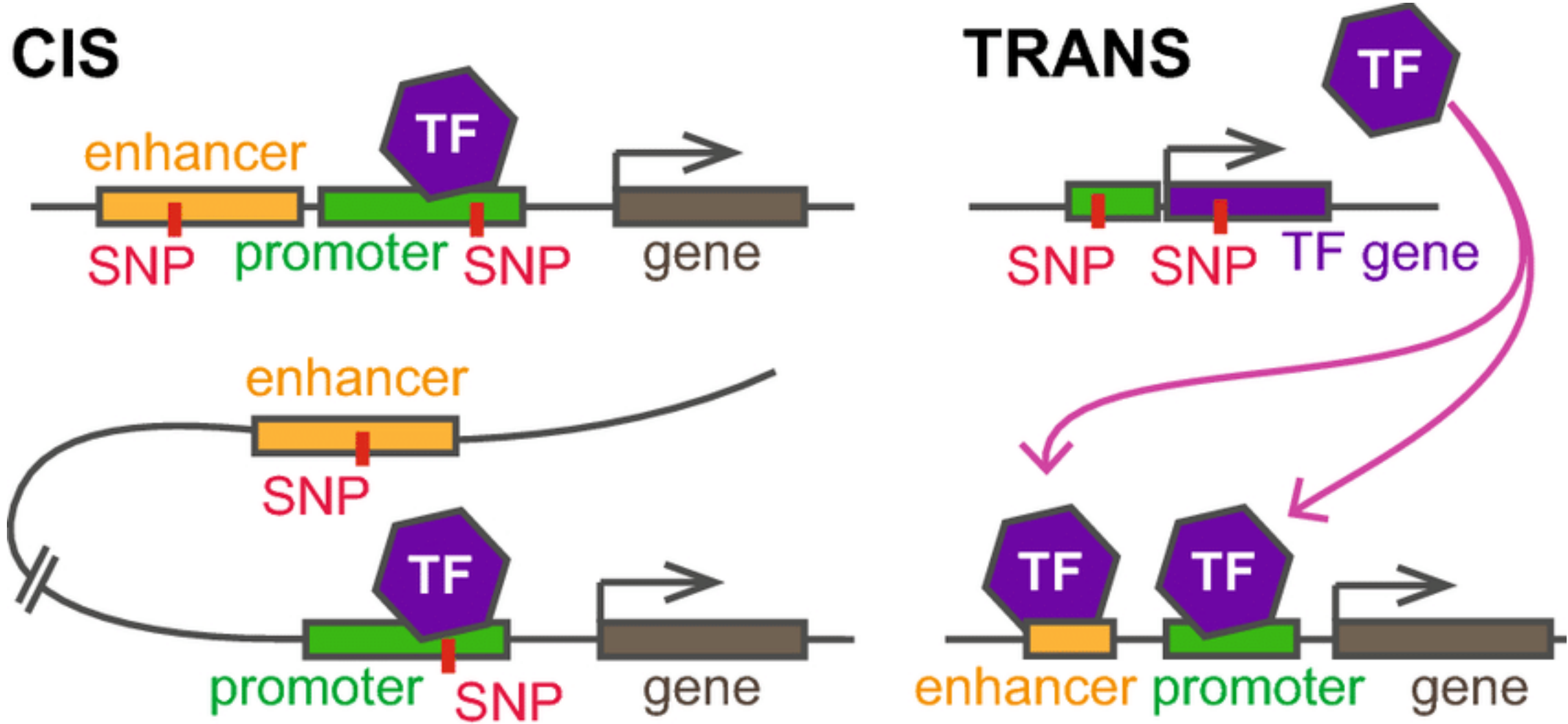
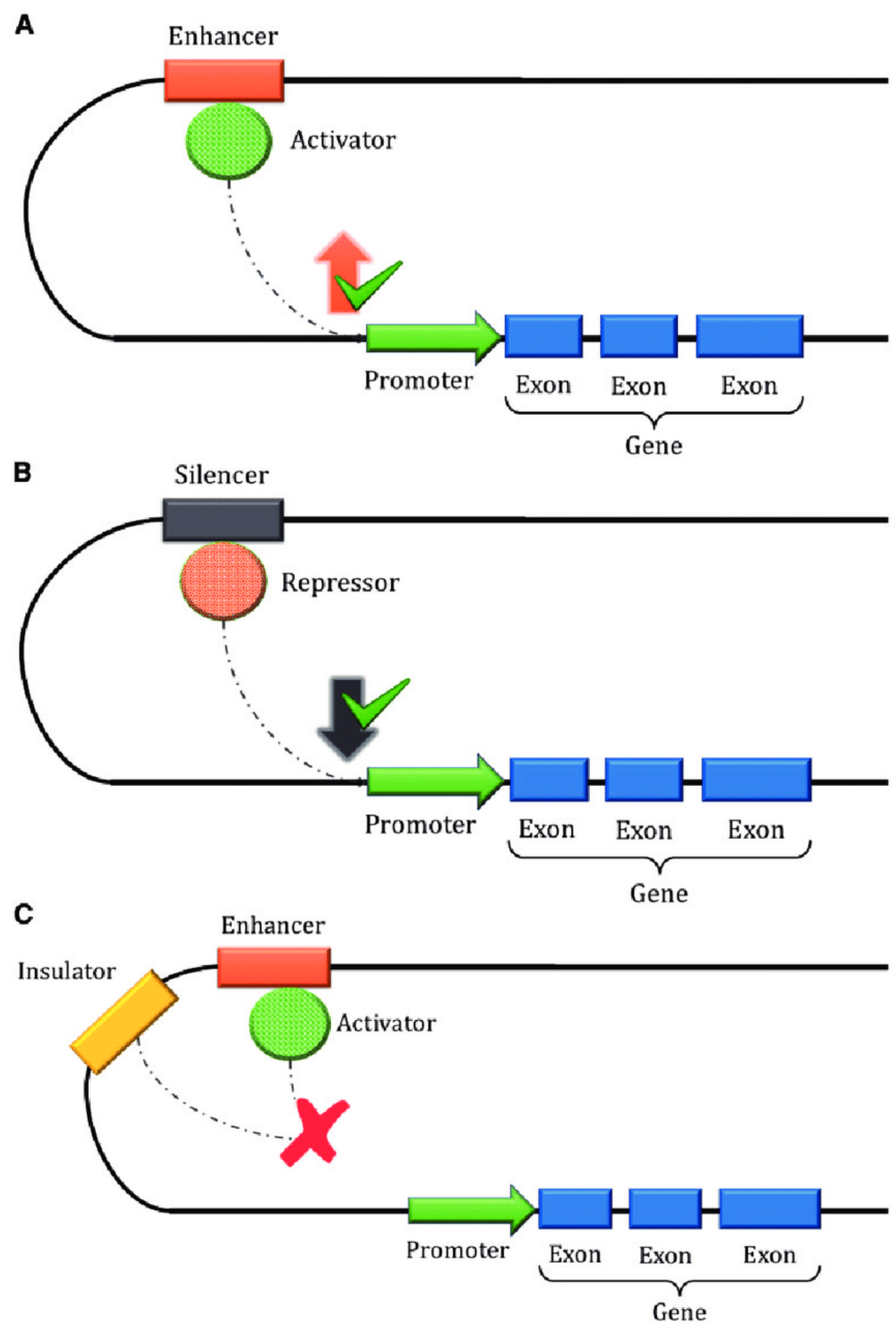
## Transcription Regulators

- A broader category that includes any molecule (protein, RNA, or other factor) that influences gene expression at any level — transcriptional, post-transcriptional, translational, or post-translational
- Includes:
  - ✓ Transcription factors (so all TFs are regulators)
  - ✓ Epigenetic modifiers (e.g., histone acetyltransferases, DNA methyltransferases)
  - ✓ Noncoding RNAs (e.g., miRNAs, lncRNAs)
  - ✓ RNA-binding proteins
  - ✓ Chromatin remodelers

# *cis* and *trans*- regulatory sequences

- **Cis-Regulatory Sequences (or Elements)** = segments of noncoding DNA located on the same molecule (“cis”) as the gene they regulate
  - They serve as binding sites for transcription factors and other proteins that control transcription
  - They control when, where, and how much a gene is transcribed
  - They include:
    - ✓ **Promoters** — right upstream of the transcription start site; where RNA polymerase and general TFs bind
    - ✓ **Enhancers** — increase transcription; can be far away from the gene (upstream, downstream, or even within introns)
    - ✓ **Silencers** — decrease transcription
    - ✓ **Insulators** — block enhancer–promoter communication or separate chromatin domain
- **Trans-Regulatory Sequences (or Elements)** = segments of noncoding DNA located on another molecule (not on the same DNA molecule as the gene being regulated)
- Trans-regulatory elements are usually genes that encode diffusible molecules — such as proteins or RNAs — that bind to *cis*-regulatory sequences on DNA to regulate transcription.

# *cis* and *trans*- regulatory sequences



# Control of transcription by DNA-binding proteins

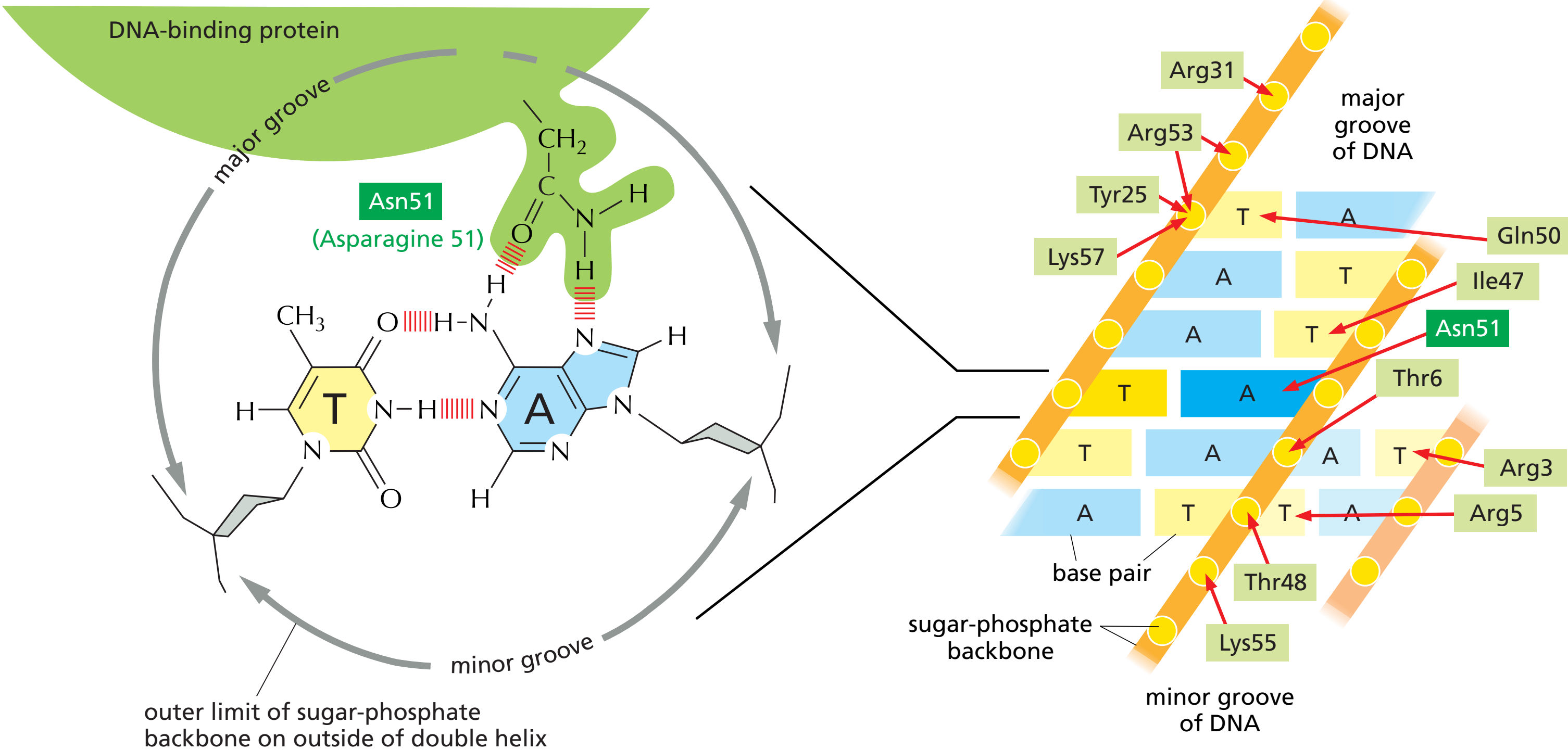
- Proteins known as **transcription factors**
- Bind **DNA sequences** (5-10 nucleotides) known as ***cis*-regulatory sequences**
- **Each gene** is controlled by its own set of *cis*-regulatory sequences (often in the intergenic region upstream of the transcription start)
- They determine the **time and place** each gene is transcribed

Transcriptional **repressors** turn genes off, transcriptional **activators** turn genes on

# Control of transcription by DNA-binding proteins

How does the **recognition** work?

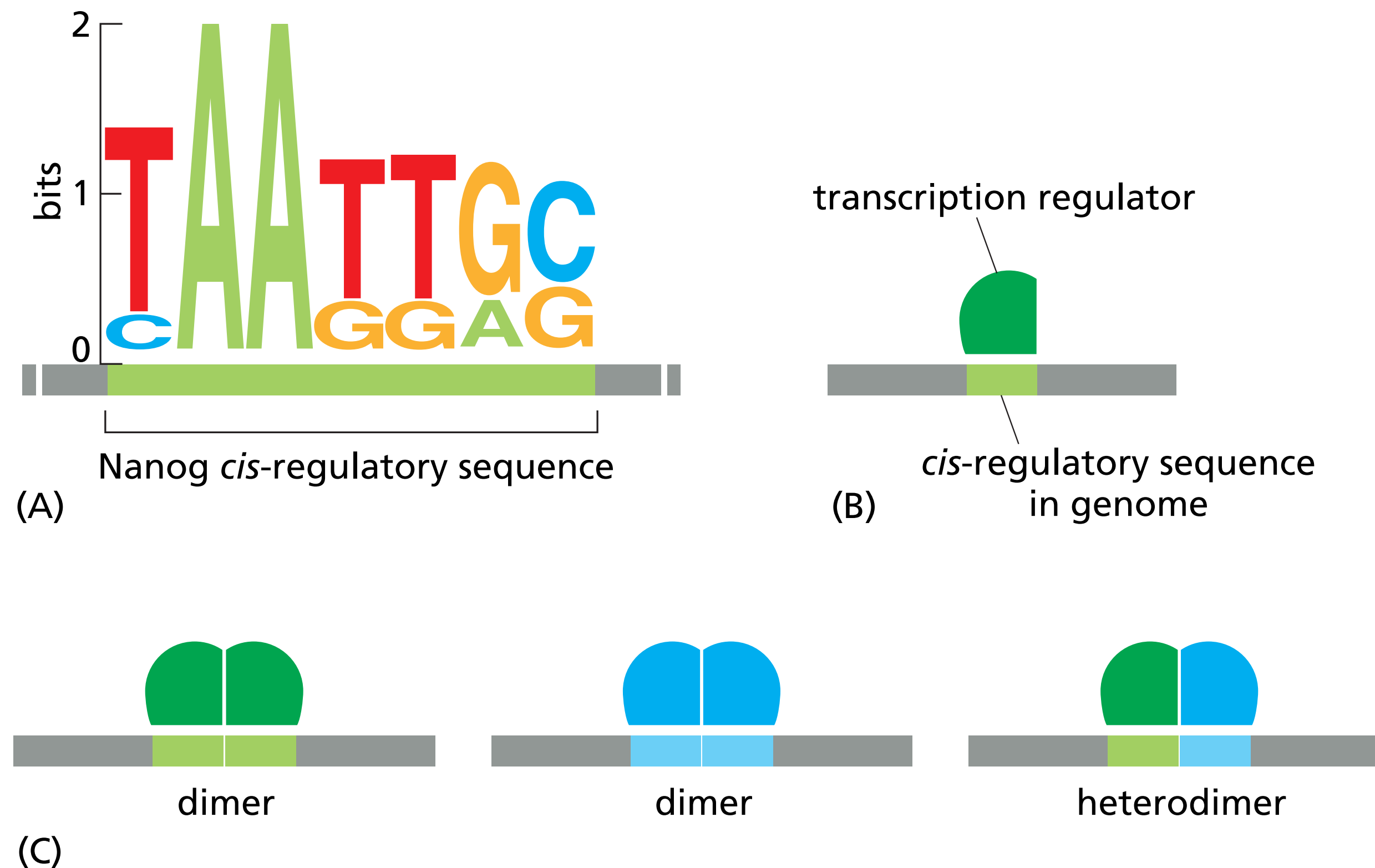
- the surface of the protein is **complementary** to the special surface features of a portion of DNA
- contacts involve **hydrogen bonds, ionic bonds and hydrophobic interactions**



# Control of transcription by DNA-binding proteins

How does the **recognition** work?

- DNA binding proteins bind to different **closely-related sequences** (depicted as a logo)
- Specificity is increased by **dimer formation** which doubles the length of the cis-sequence



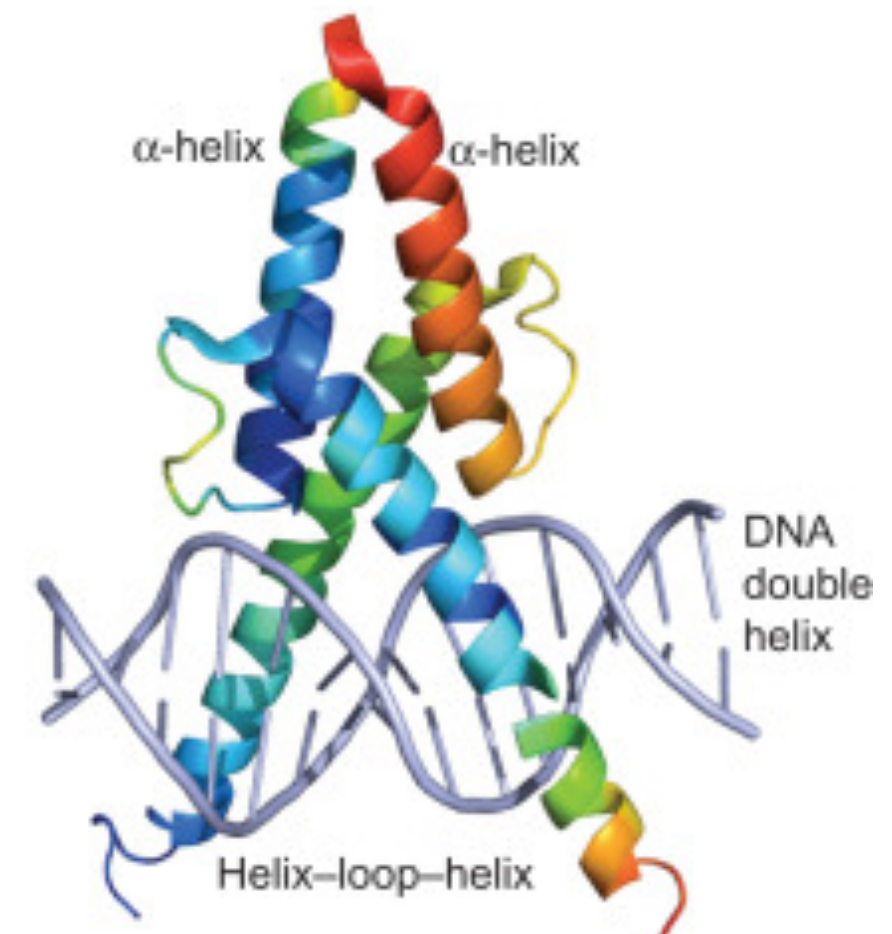
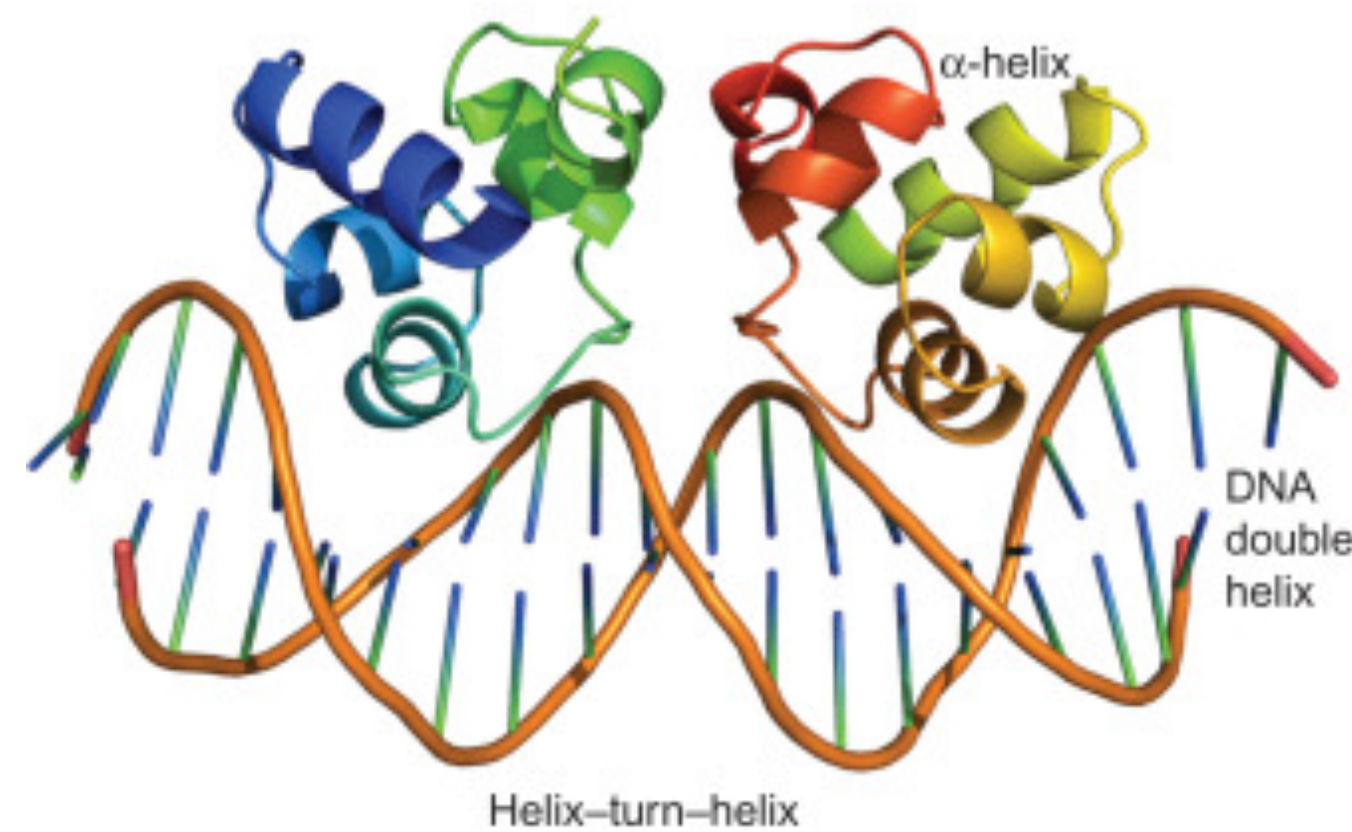
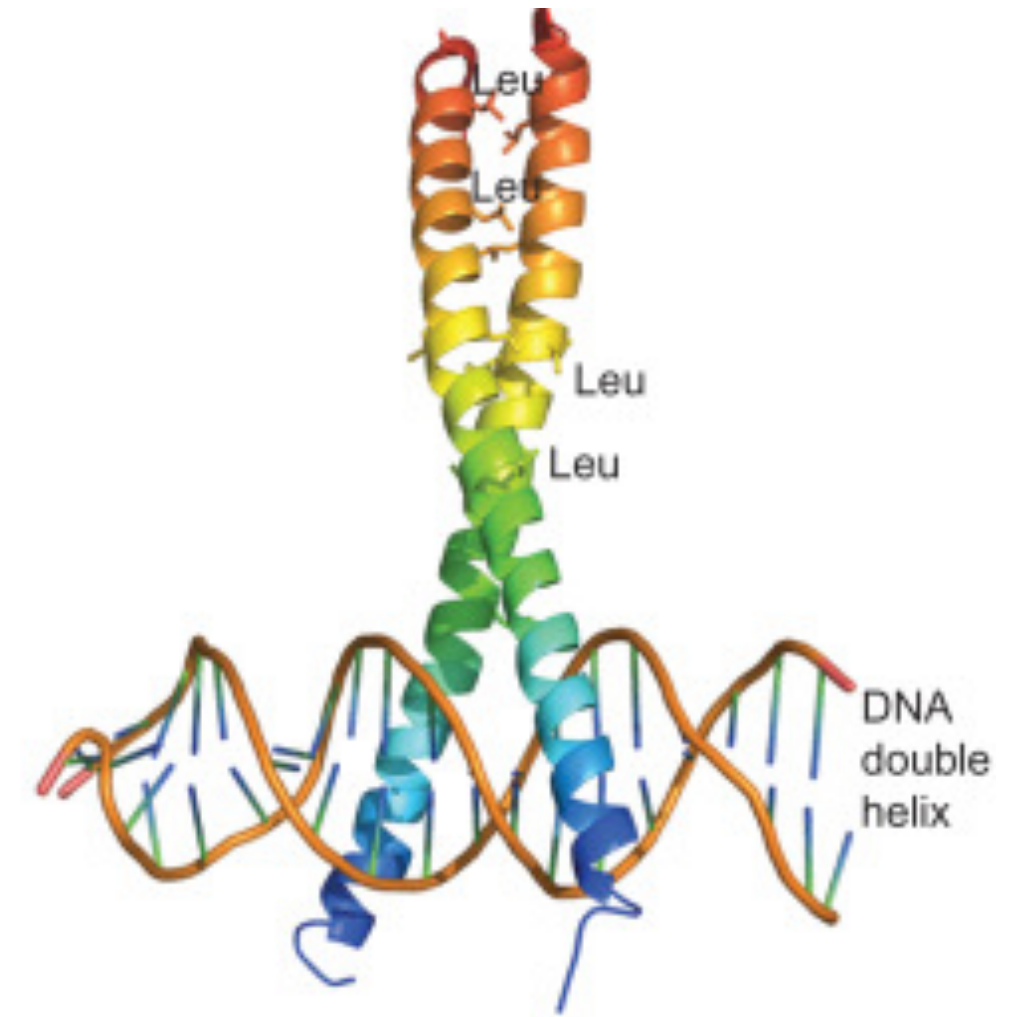
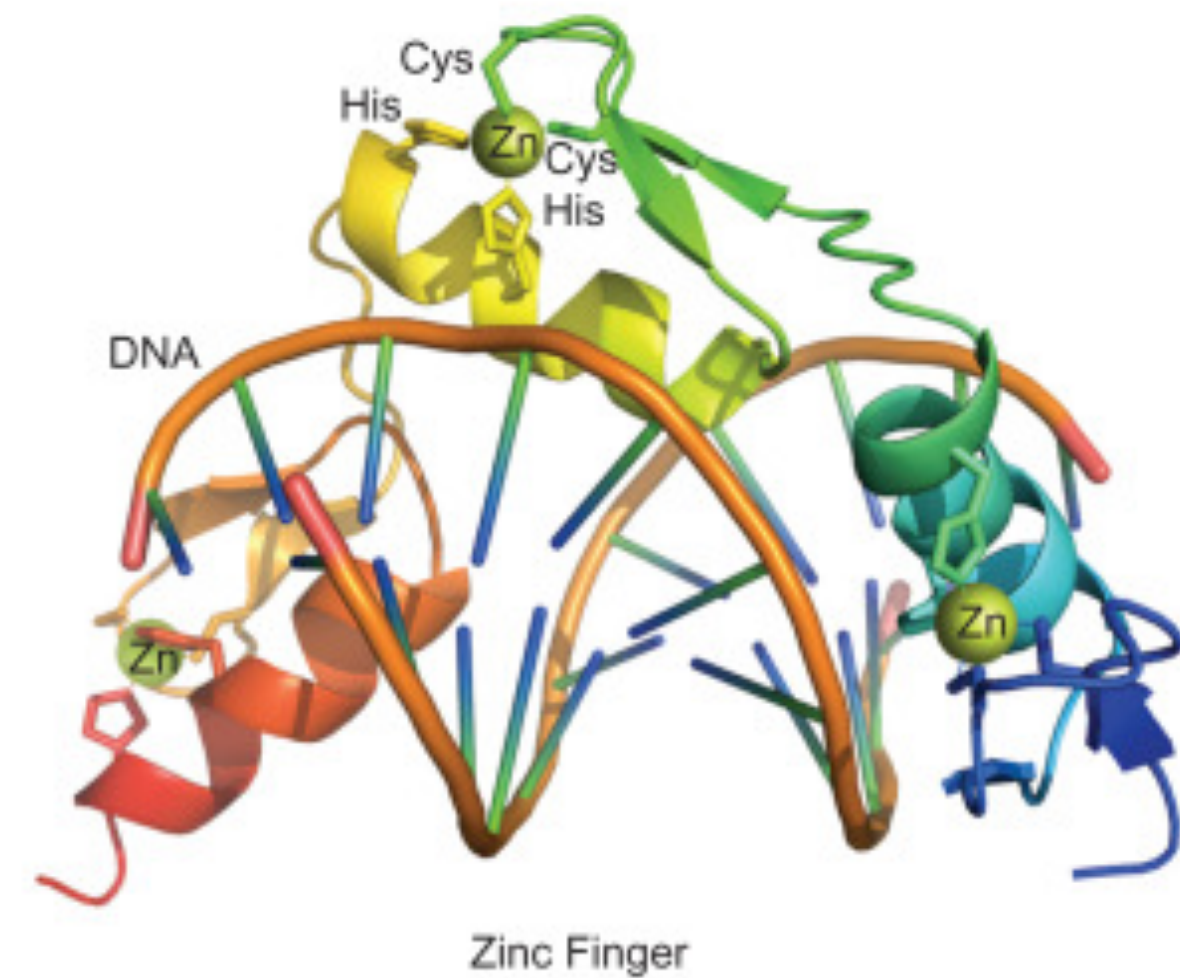
# Control of transcription by DNA-binding proteins

## General Structure of Transcription Factors

They typically have distinct **domains**, each performing a specific function.

### 1. DNA-Binding Domain (DBD)

- Recognizes and binds to specific cis-regulatory DNA sequences (like promoters or enhancers)
- Determines which genes the TF can regulate
- Examples of DBD motifs:
  - Helix-Turn-Helix (HTH)
  - Zinc Finger
  - Leucine Zipper (bZIP)
  - Helix-Loop-Helix (bHLH)
  - Homeodomain



# Control of transcription by DNA-binding proteins

## 2. Transactivation (or Repression) Domain

- Interacts with other proteins (like coactivators, repressors, or components of the transcription machinery).
- Activators use this domain to recruit:
  - Mediator complex
  - Histone acetyltransferases (HATs)
  - Chromatin remodelers
- Repressors use it to recruit:
  - Histone deacetylases (HDACs)
  - Corepressor complexes

## 3. Dimerization Domain

- Many TFs function as dimers (either homo- or heterodimers).
- Enables two TF molecules to pair up, which increases binding specificity and stability.

## 4. Ligand-Binding or Regulatory Domain (in some TFs)

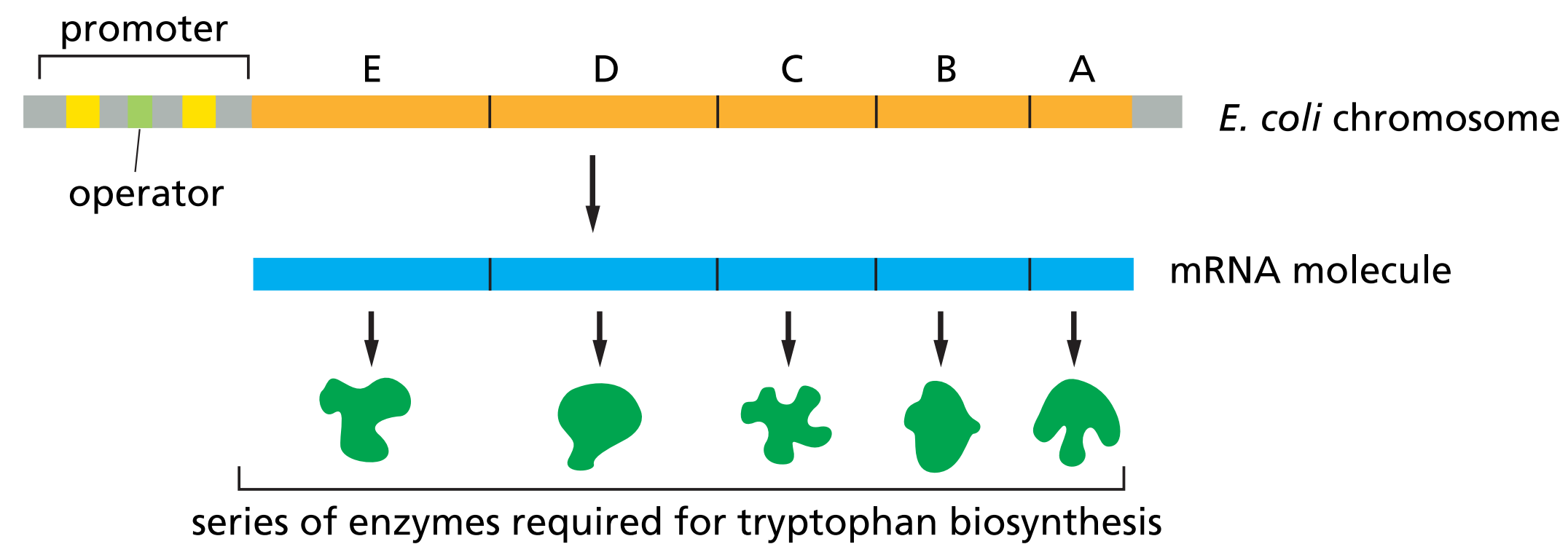
- Found in signal-responsive TFs like nuclear hormone receptors

# Control of transcription by DNA-binding proteins

Transcription factors **switch genes on and off**

- Example: the tryptophan repressor (bacteria)

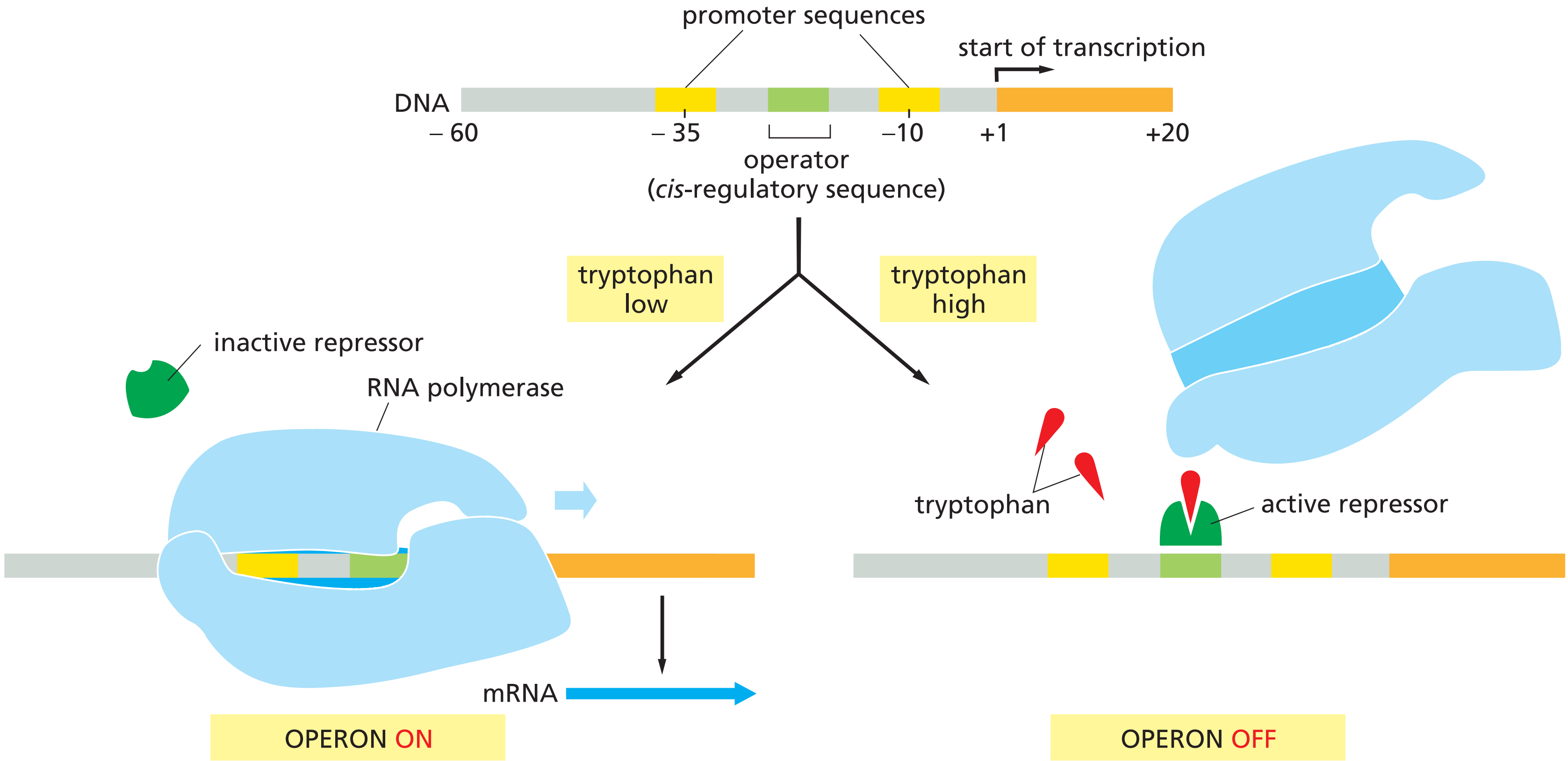
The tryptophan operon



- When tryptophan levels are low, the operon is transcribed
- When tryptophan levels are high, the operon is not transcribed

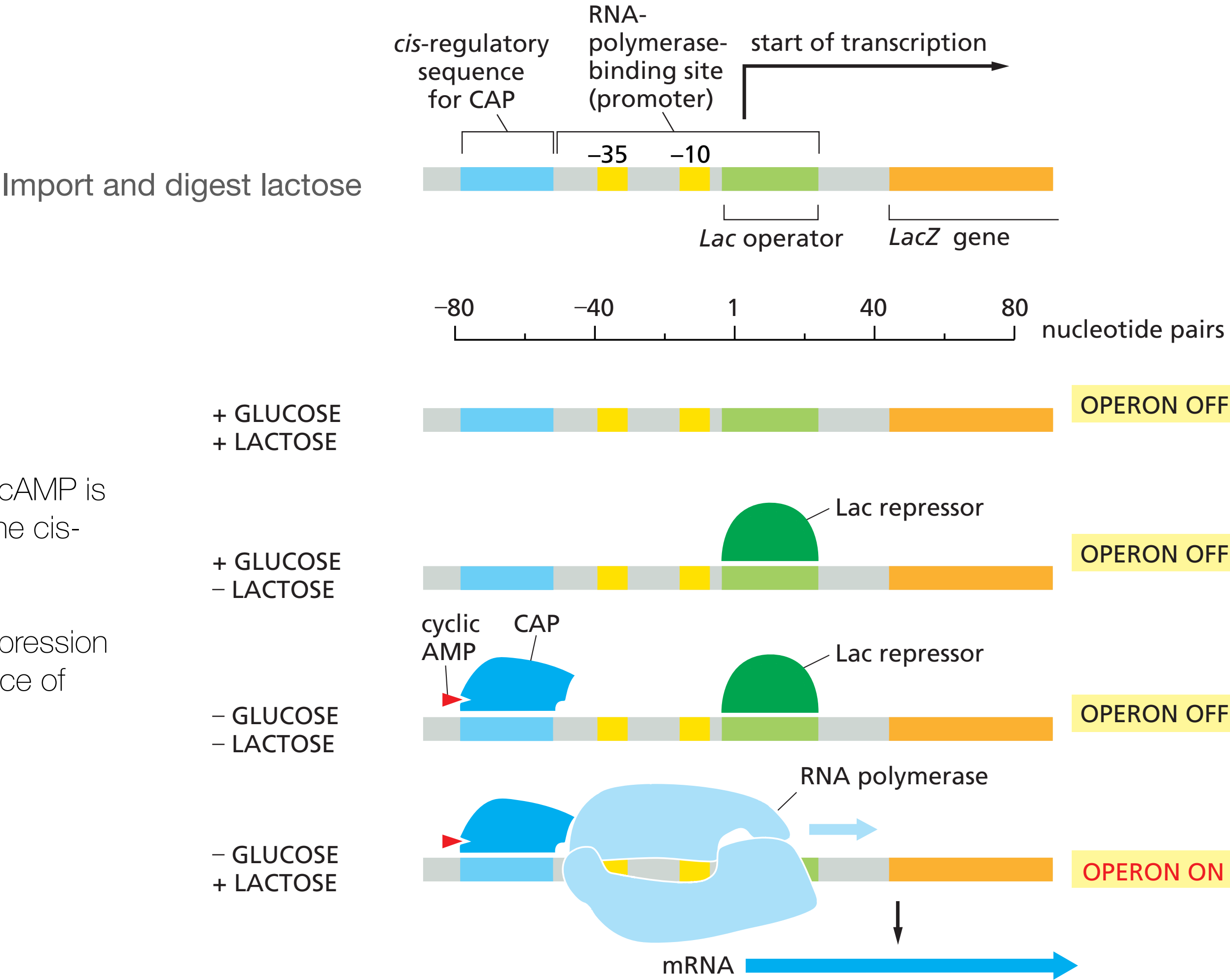
# Control of transcription by DNA-binding proteins

- Example: The tryptophan repressor (bacteria)



# Control of transcription by DNA-binding proteins

- Example: The lac operon (bacteria)

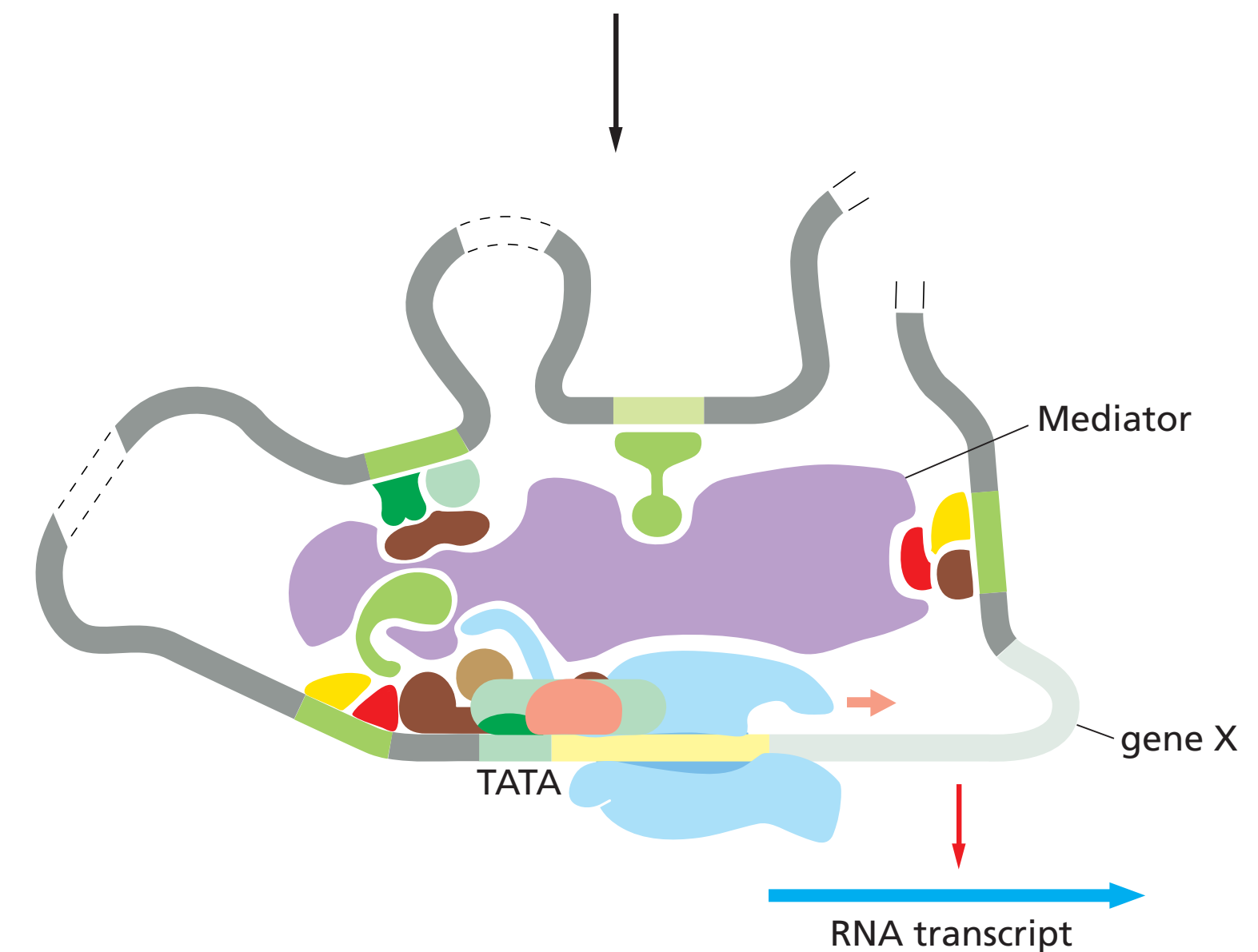
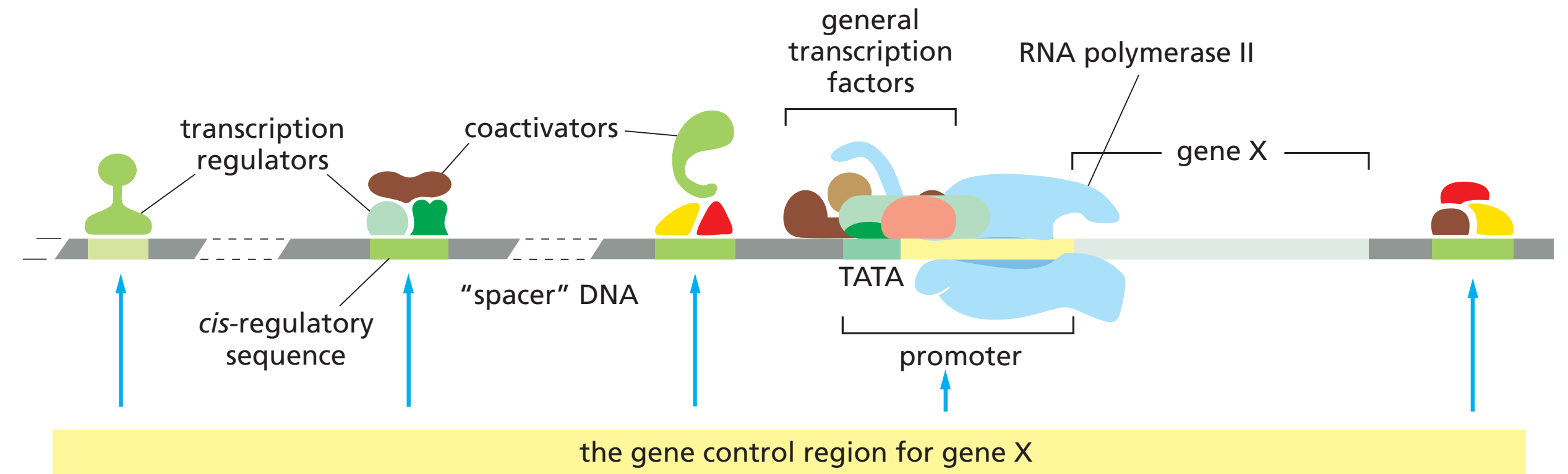


When glucose levels are low, cAMP is produced and CAP binds to the *cis*-regulatory sequence.

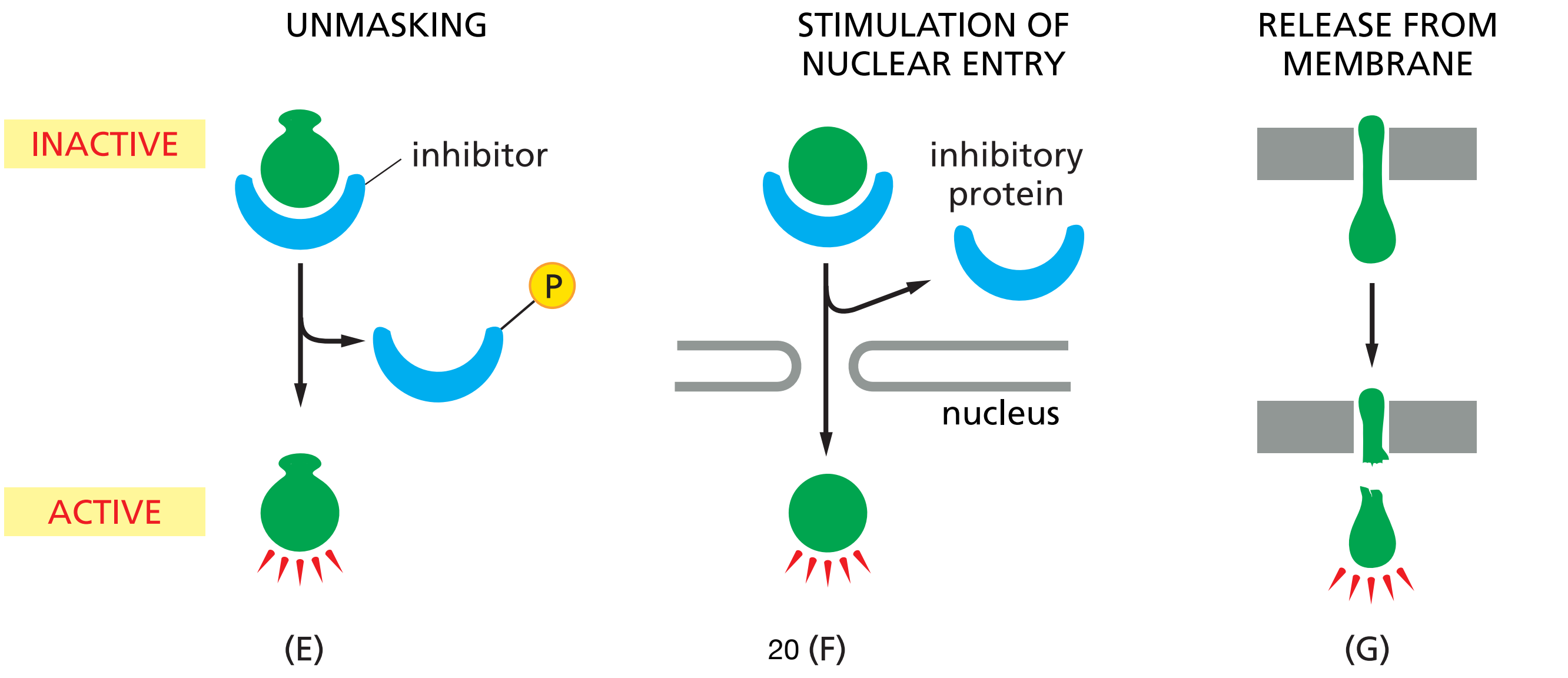
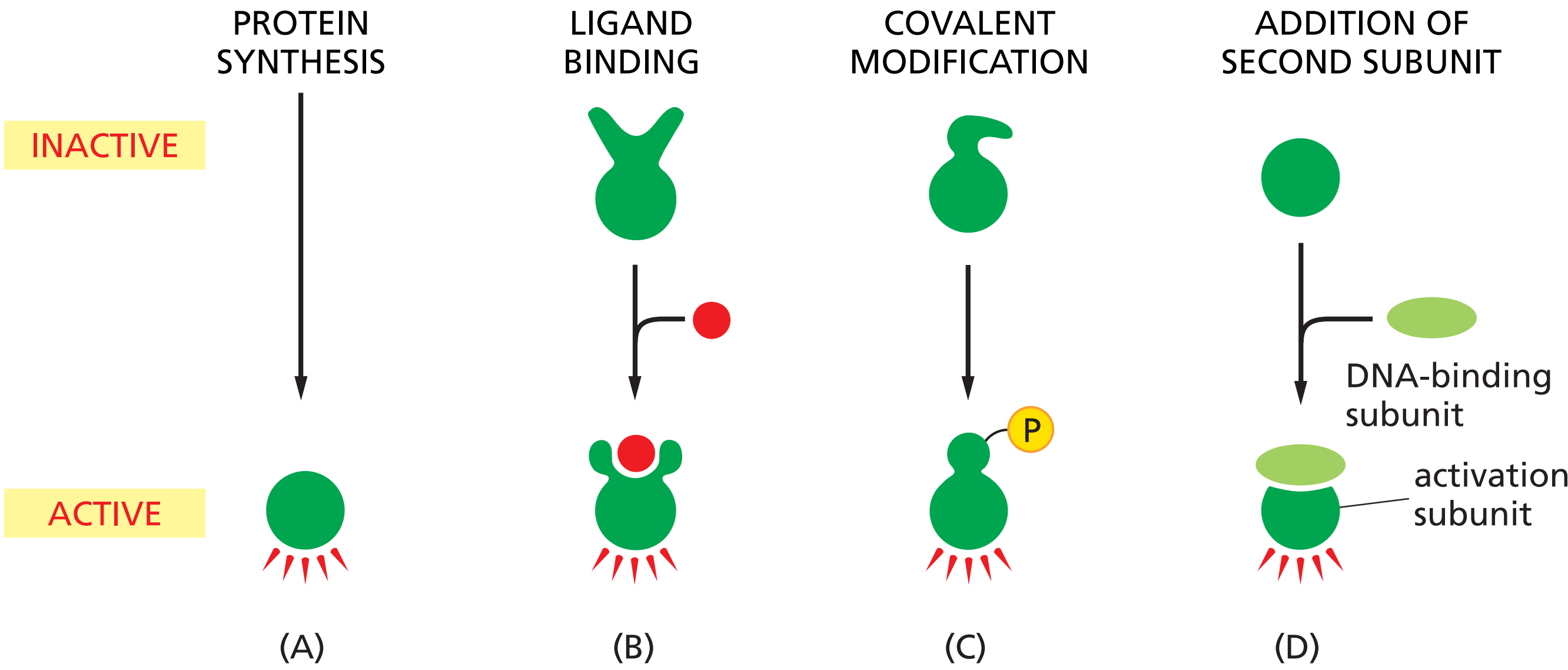
Lac repressor prevents the expression of the lac operon in the absence of lactose.

# Control of transcription by DNA-binding proteins

- More complex in **Eukaryotes**
- RNA polymerase II requires **5 general transcription factors** - their stepwise assembly already controls the rate of transcription initiation
- Cis-regulatory sequences spread over a long stretch of DNA = **gene control region**
- The gene control region includes the **promoter + all cis-regulatory sequences**
- There are **thousands of different transcription regulators**



# Modulation of the activity of transcription regulators



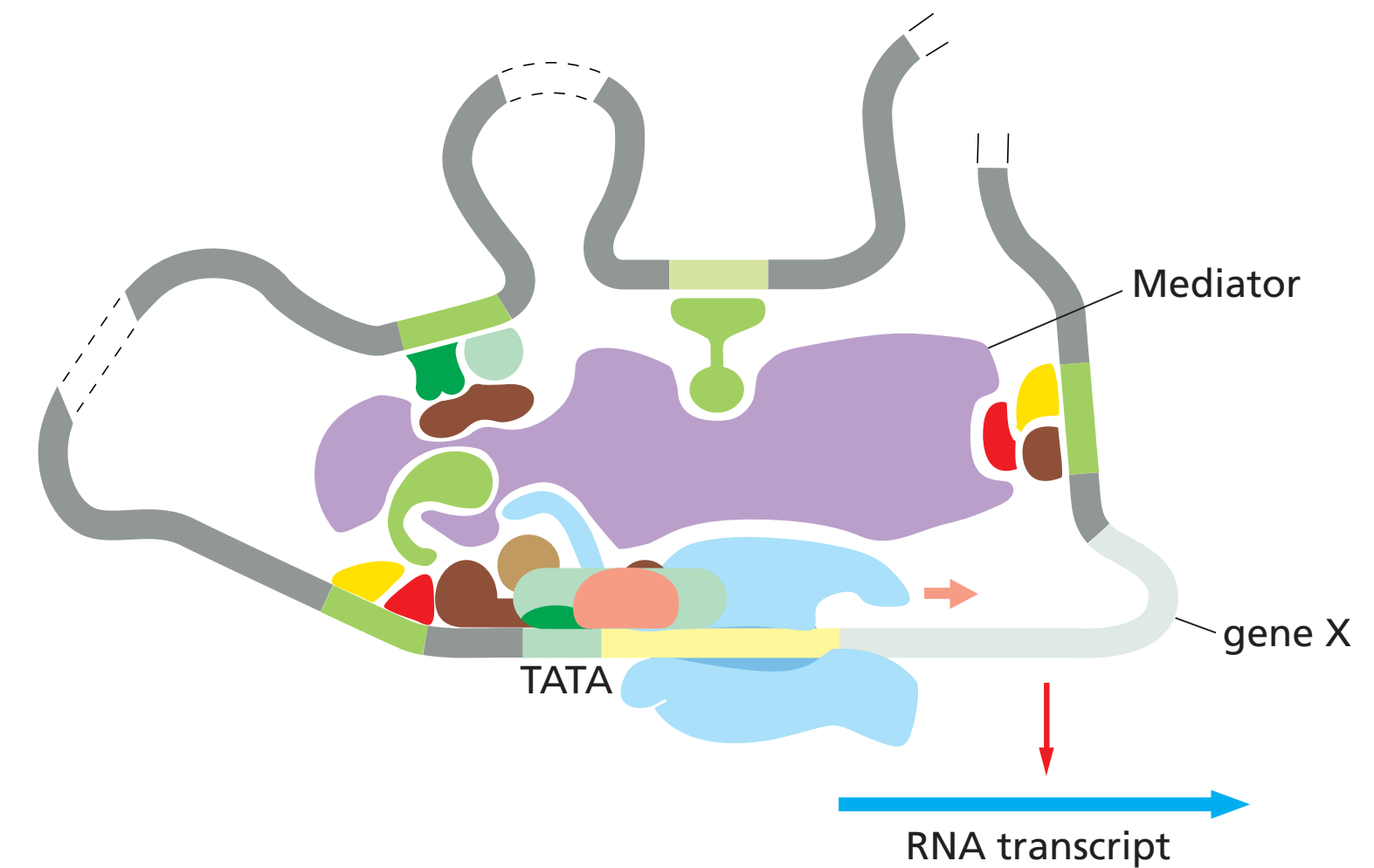
# IV. Control of gene expression

## 1. Transcriptional control

- a. Transcription regulators (**activators**, repressors)
- b. Understanding different control systems
- c. Combinatorial gene control and cell types

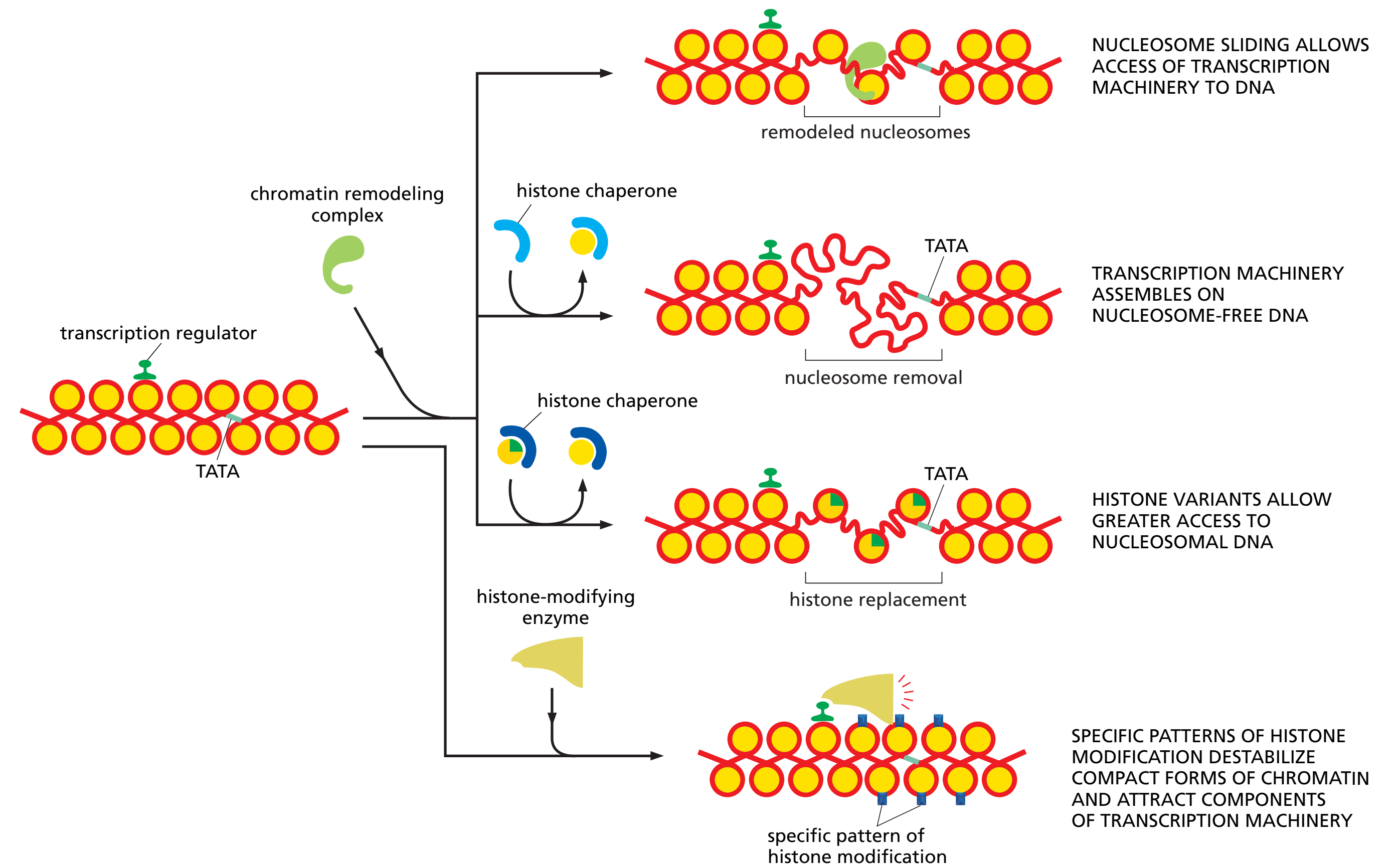
# Activator proteins

- cis-regulatory sequences to which activator bind were called **enhancers**
- they can be **far** from the promoter (but always on the same chromosome)
- activators **attract and position** the RNA polymerase II and **release** it so transcription can begin
- the **mediator** protein is composed of 30 subunits that form a bridge between transcription activators, RNA polymerase and general transcription factors



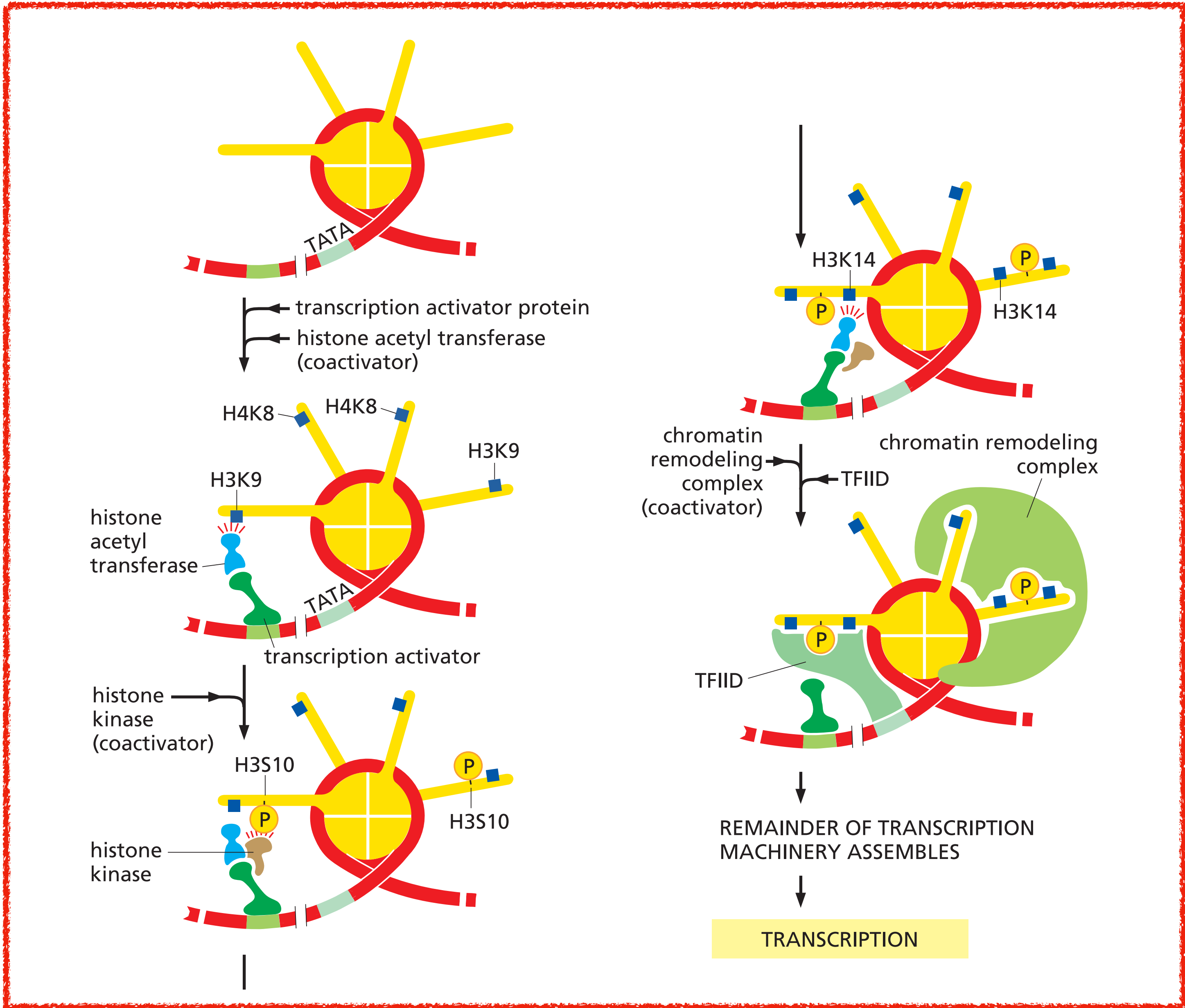
# Activator proteins

- RNA polymerase and general transcription factors are not able to assemble on a promoter that is **packaged into chromatin**
- Activators trigger changes in **chromatin structure** making DNA more accessible
- To do this they use **histone modifications, nucleosome remodelling, nucleosome removal and histone replacement** (co-activators)



# Activator proteins

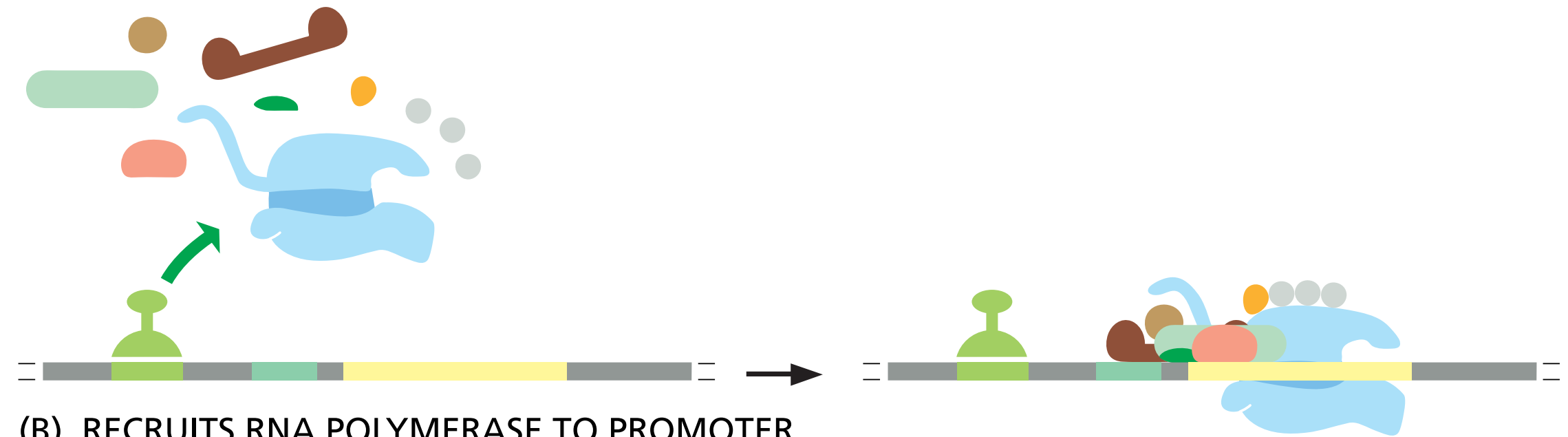
- These mechanisms work together during **transcription initiation**
- The altered chromatin can be **rapidly reversed** (quick on-off switches) or **not**



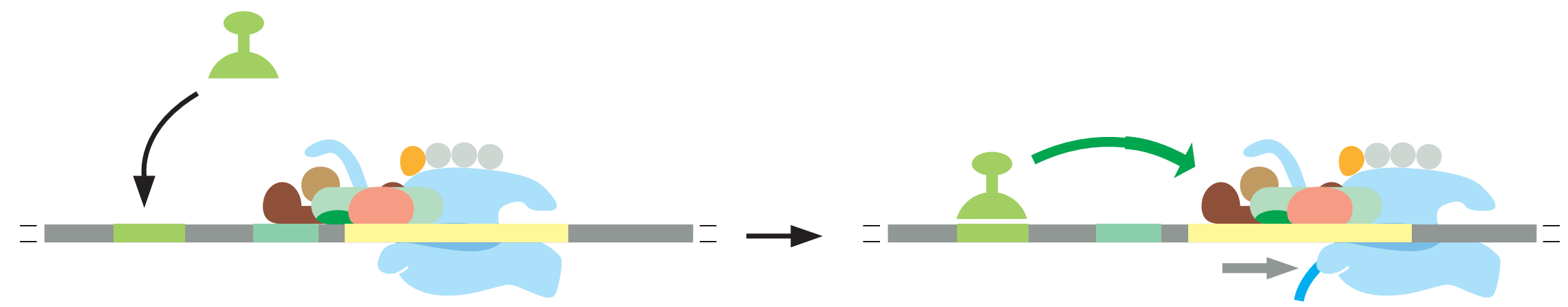
# Activator proteins



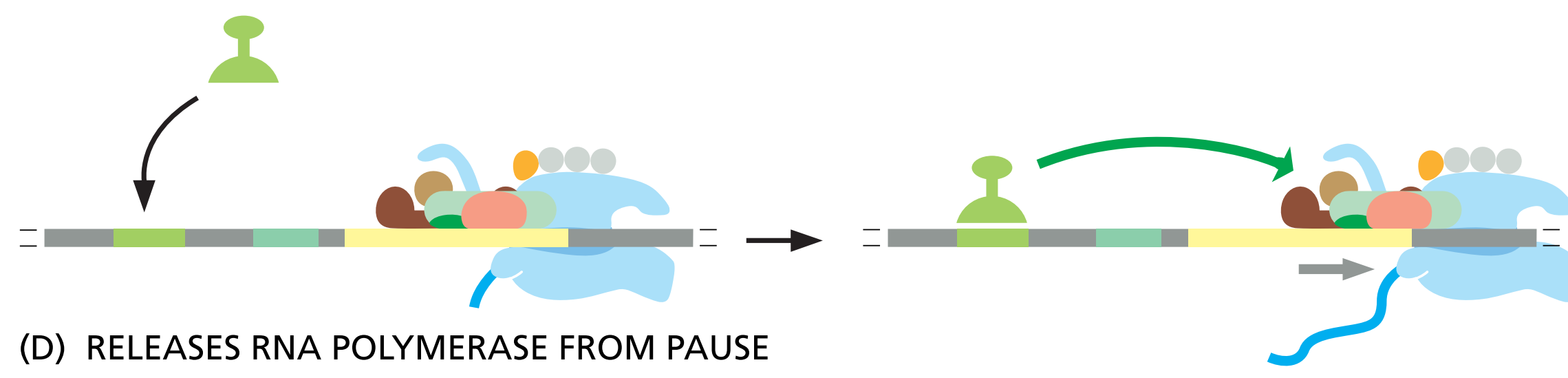
(A) PROMOTES BINDING OF ADDITIONAL REGULATORS



(B) RECRUITS RNA POLYMERASE TO PROMOTER



(C) RELEASES RNA POLYMERASE TO BEGIN TRANSCRIPTION



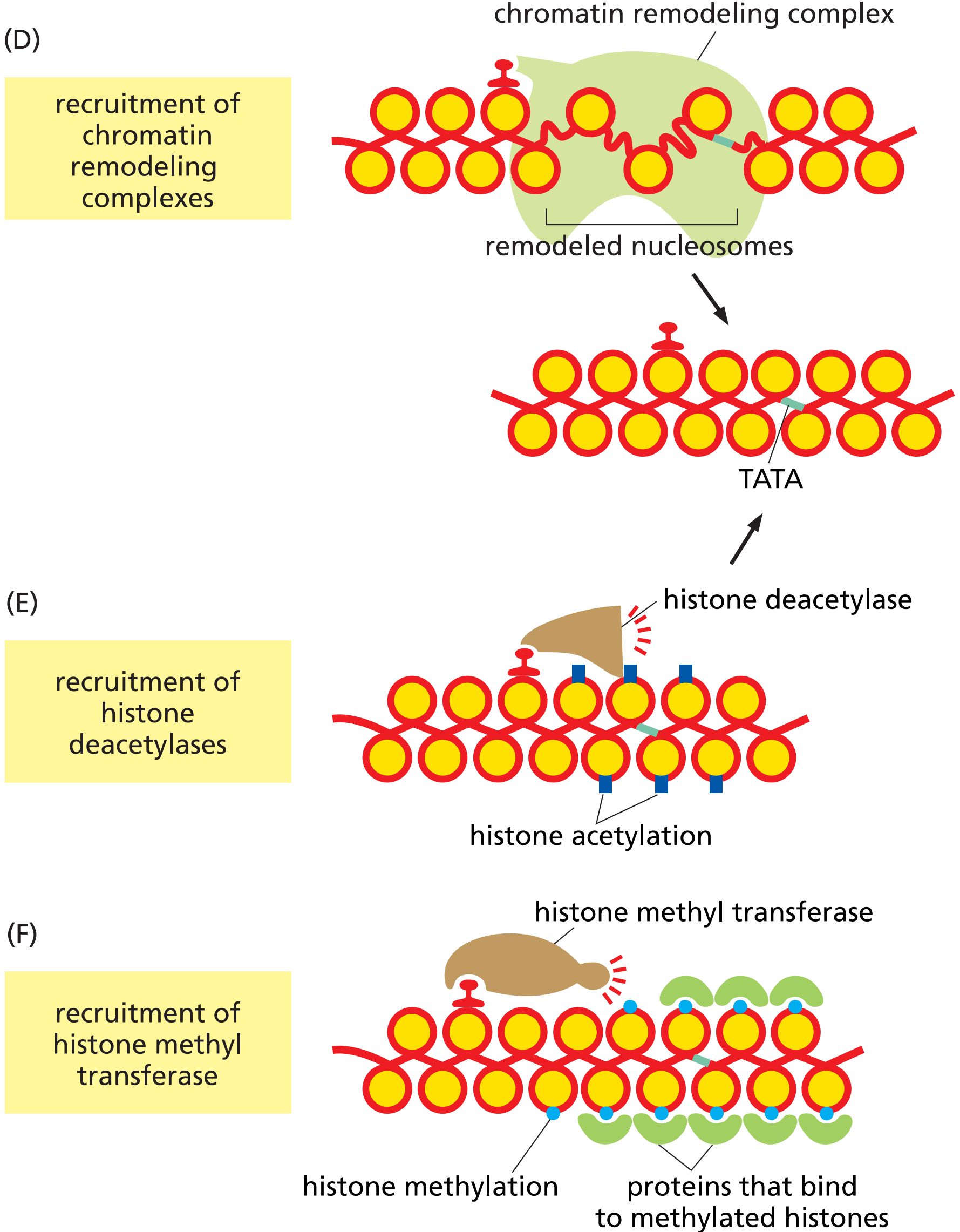
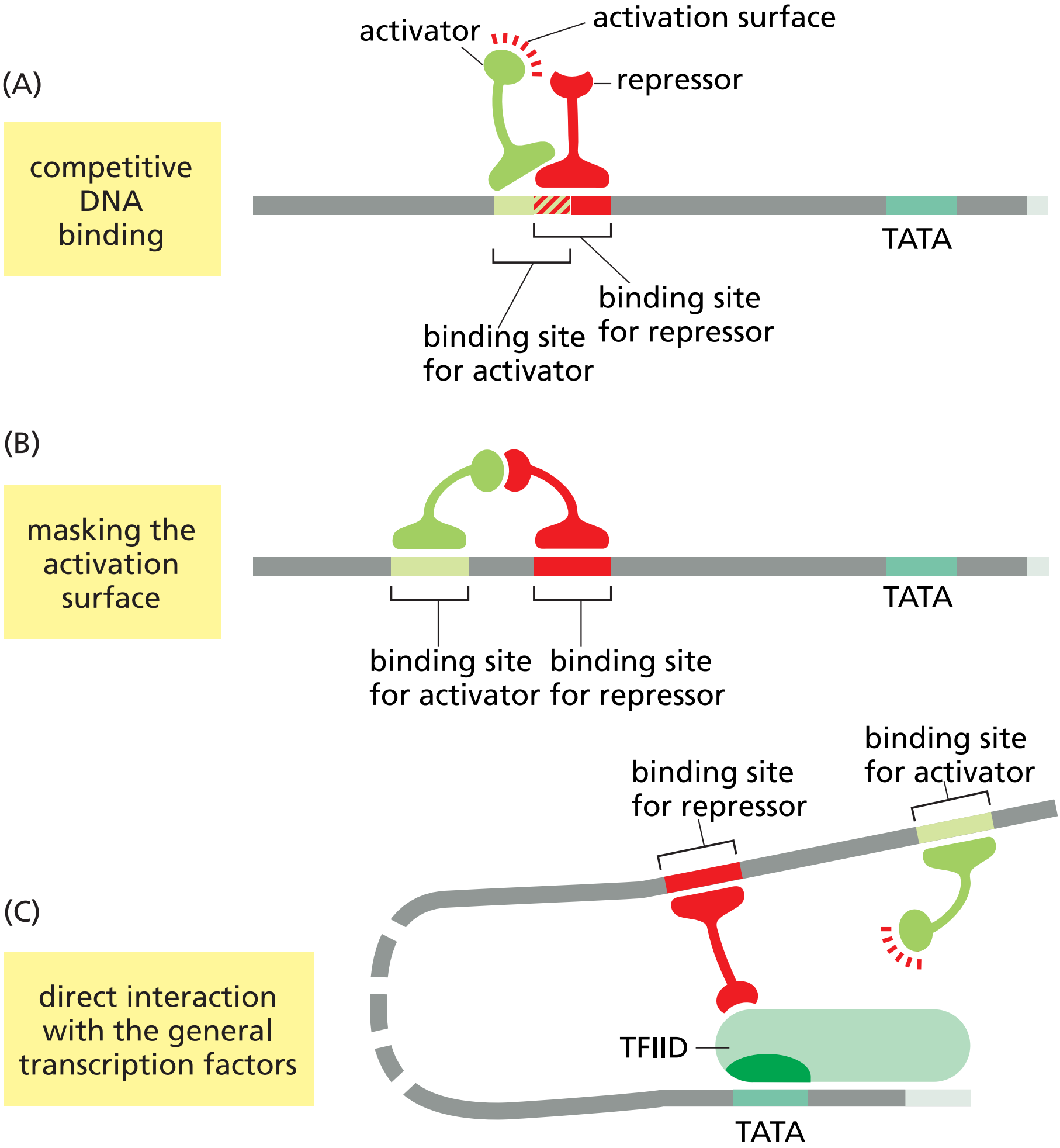
(D) RELEASES RNA POLYMERASE FROM PAUSE

# IV. Control of gene expression

## 1. Transcriptional control

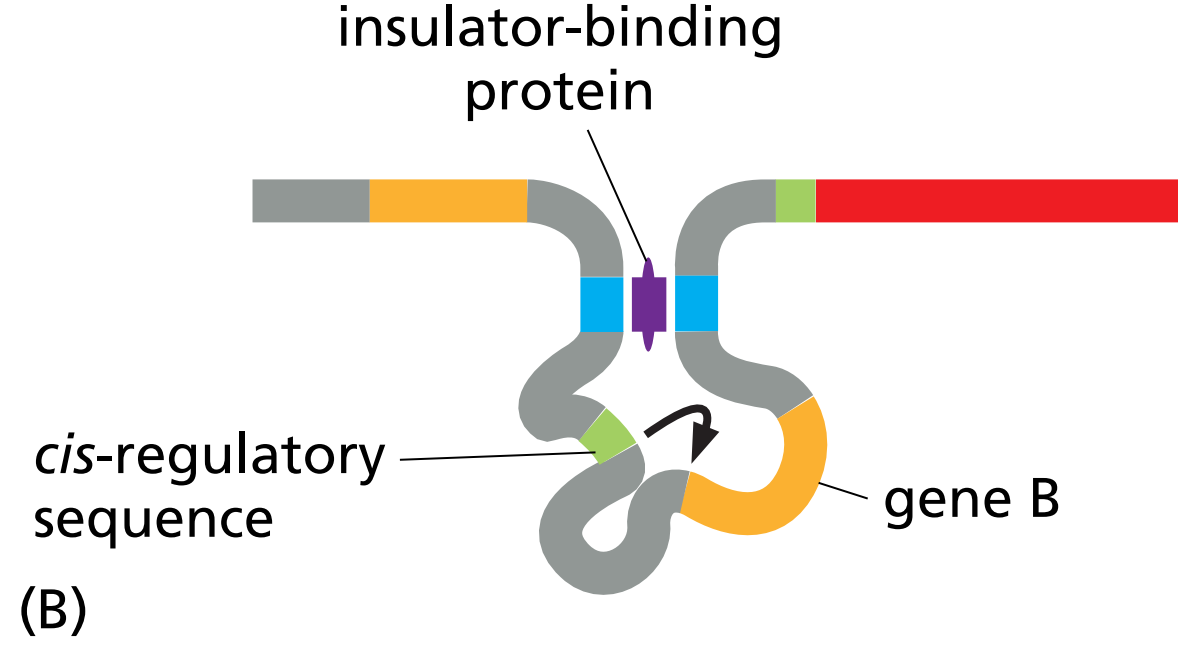
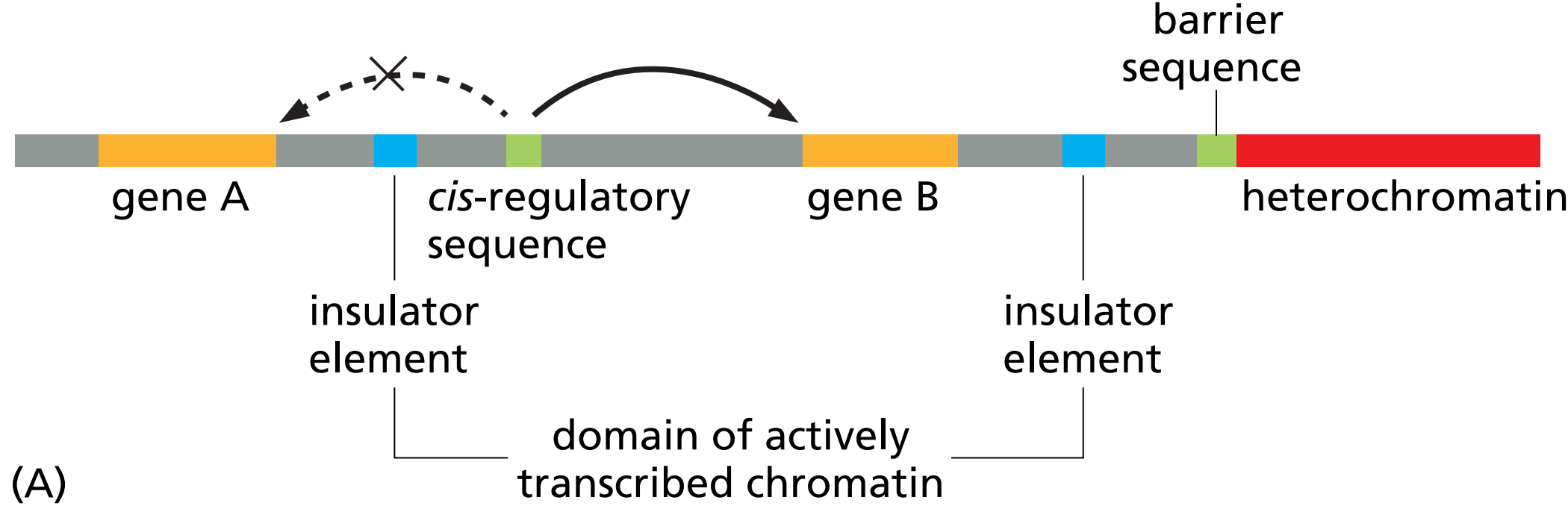
- a. Transcription regulators (activators, **repressors**)
- b. Understanding different control systems
- c. Combinatorial gene control and cell types

# Repressor proteins



# Insulator DNA sequences

- Prevent cis-regulatory elements to **activate inappropriate** genes



# IV. Control of gene expression

## 1. Transcriptional control

a. Transcription regulators (activators, repressors)

**b. Understanding different control systems**

c. Combinatorial gene control and cell types

a. Tryptophan repressor (bacteria)

b. The lac operon (bacteria)

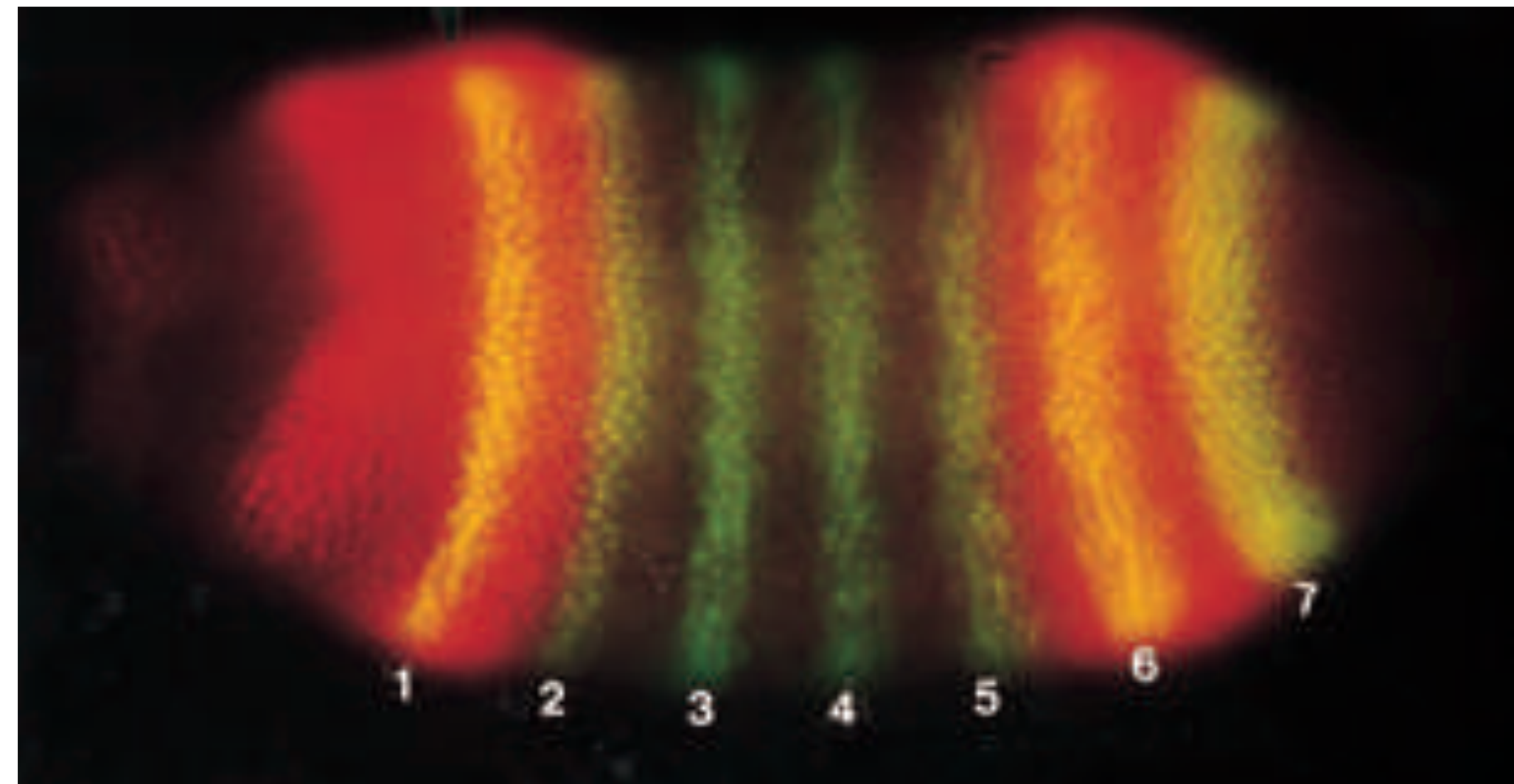
**c. The drosophila even-skipped gene**

d. The Gal4-UAS from yeast

e. The Arabinose operon (bacteria)

# Drosophila Even-skipped gene

- Plays an important role in Drosophila embryo
- The embryo is a single giant cell with multiple nuclei (~6000) in a giant cytoplasm with a mixture of transcription factors
- Those are distributed unevenly
- The nuclei rapidly start to express different genes because they are exposed to different transcription regulators

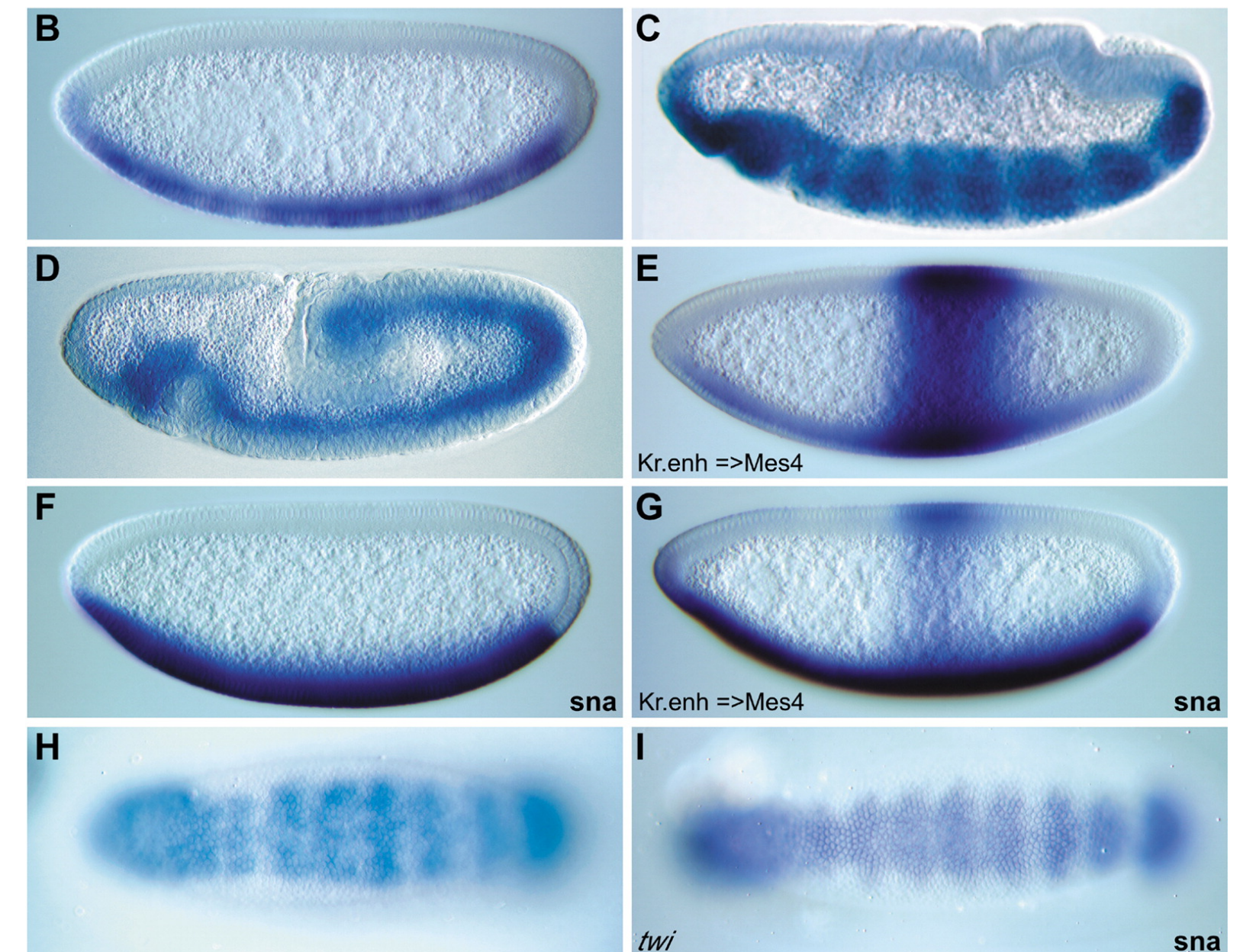
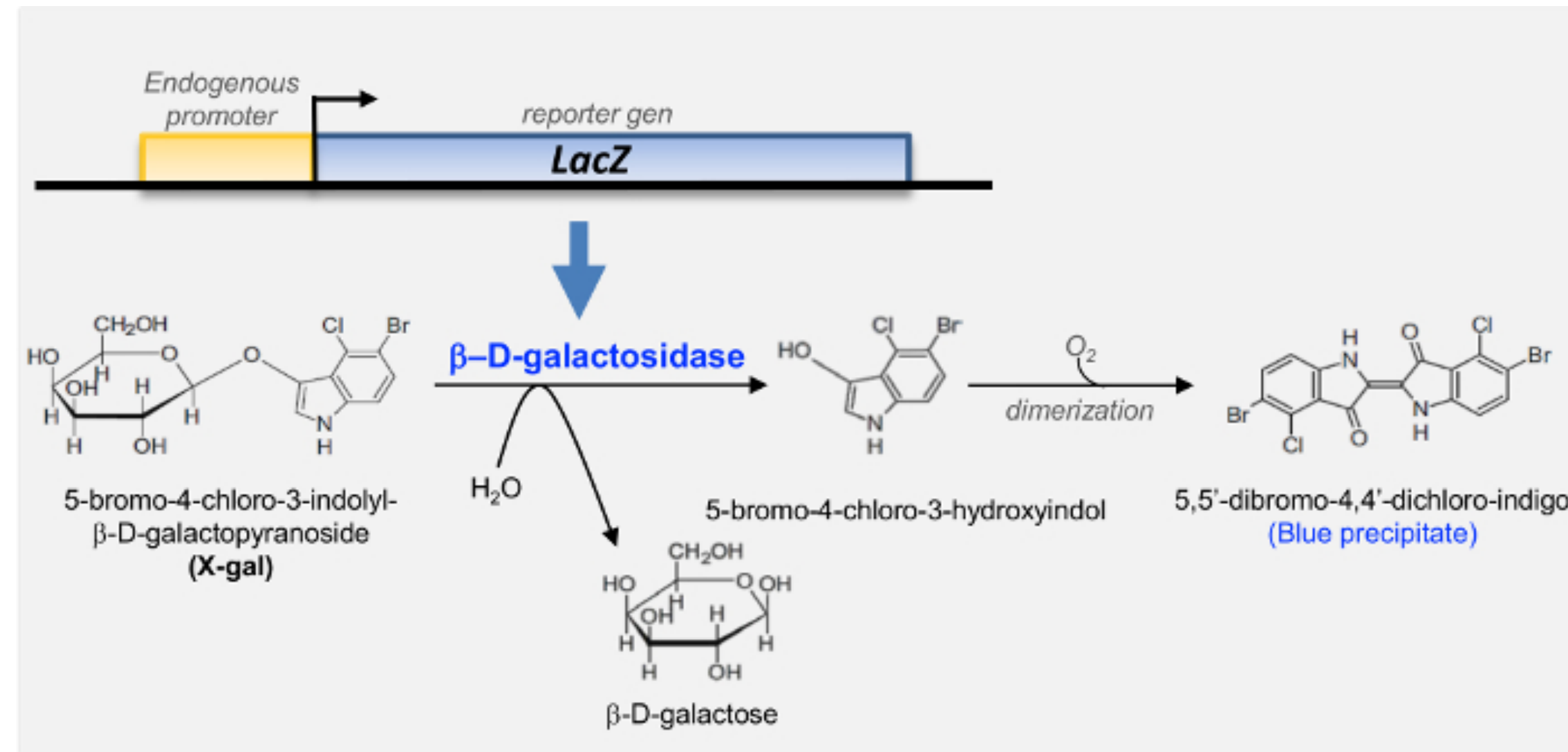
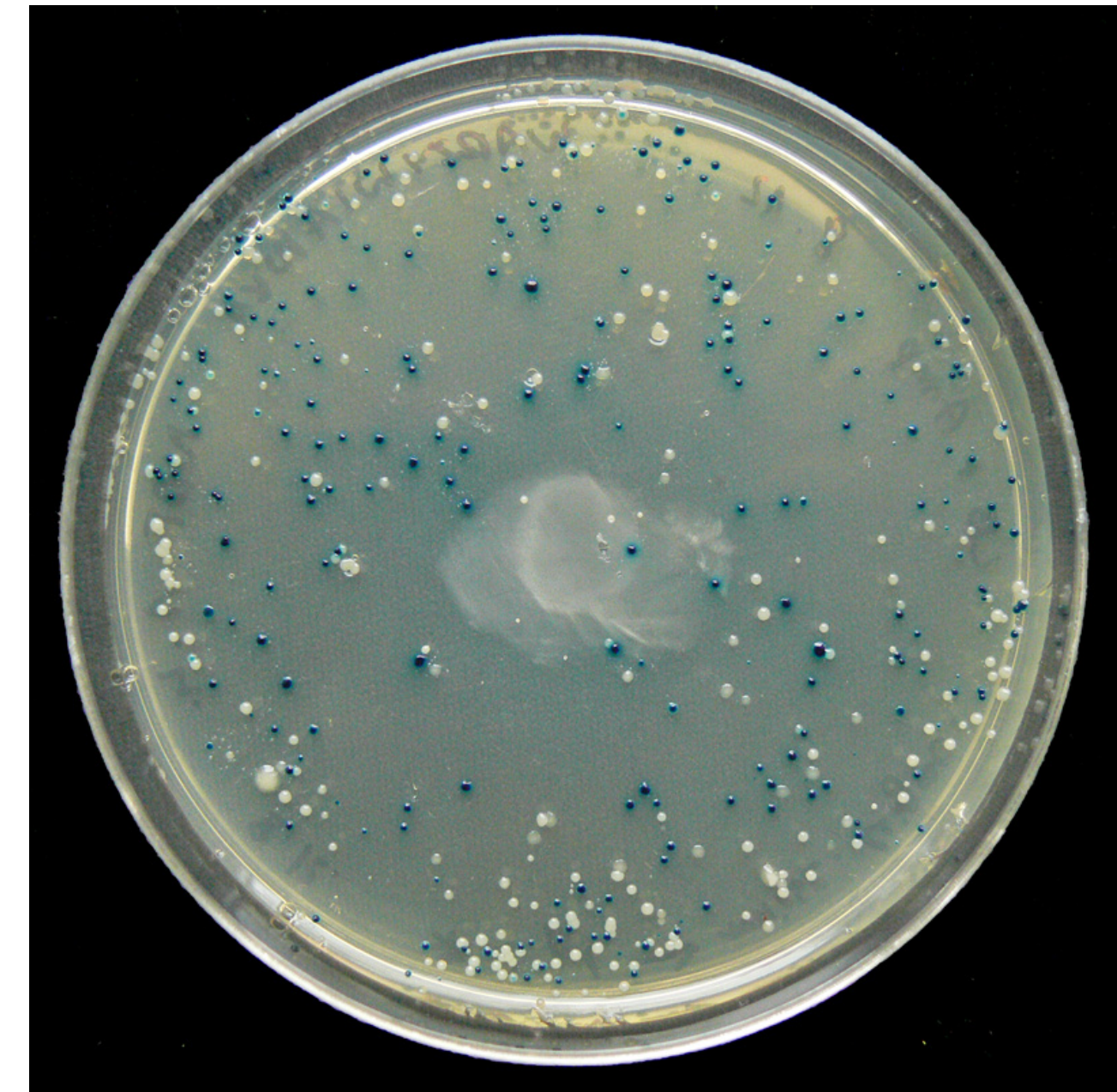


Eve is expressed in 7 precisely positioned stripes



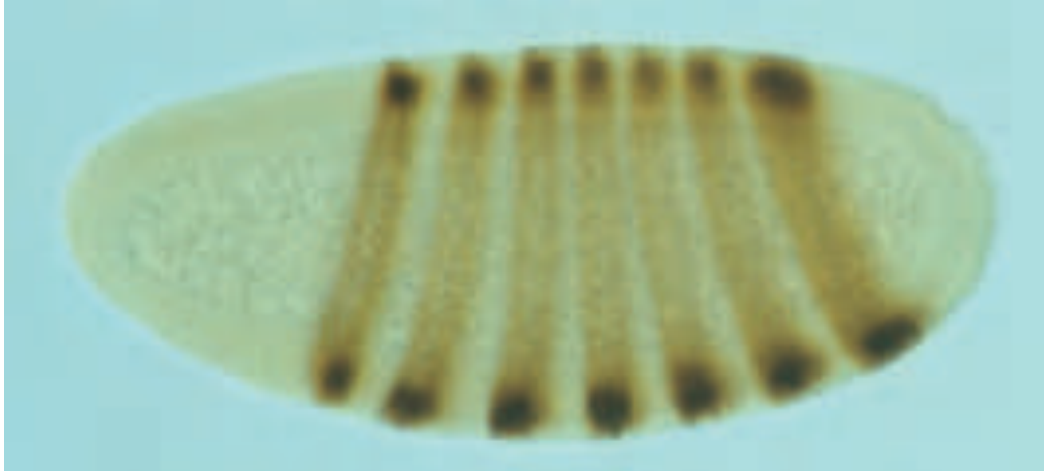
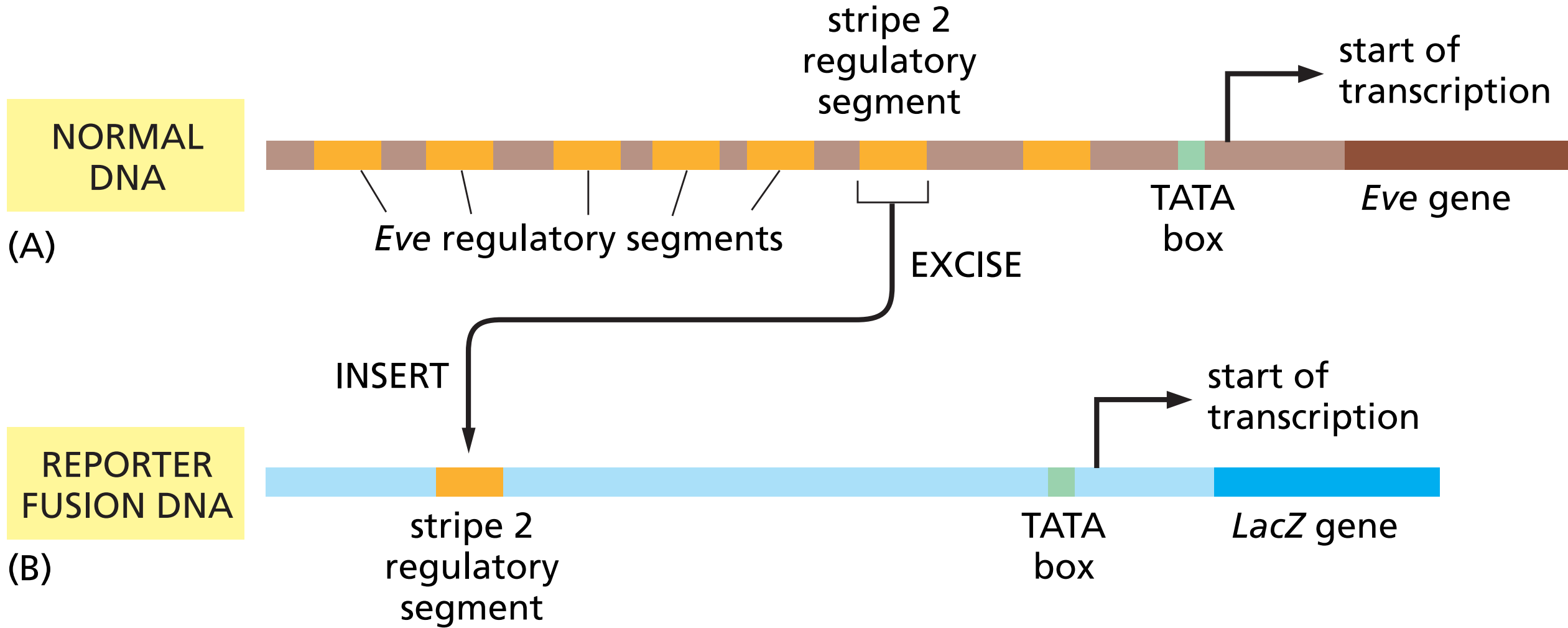
# Drosophila Even-skipped gene

- Reporter genes: a **reporter gene** is a gene that scientists attach to another gene or regulatory sequence in order to **track or measure gene expression**.
- Ex: *lacZ*



# Drosophila Even-skipped gene

- How did we link the DNA regulatory sequence to its function?



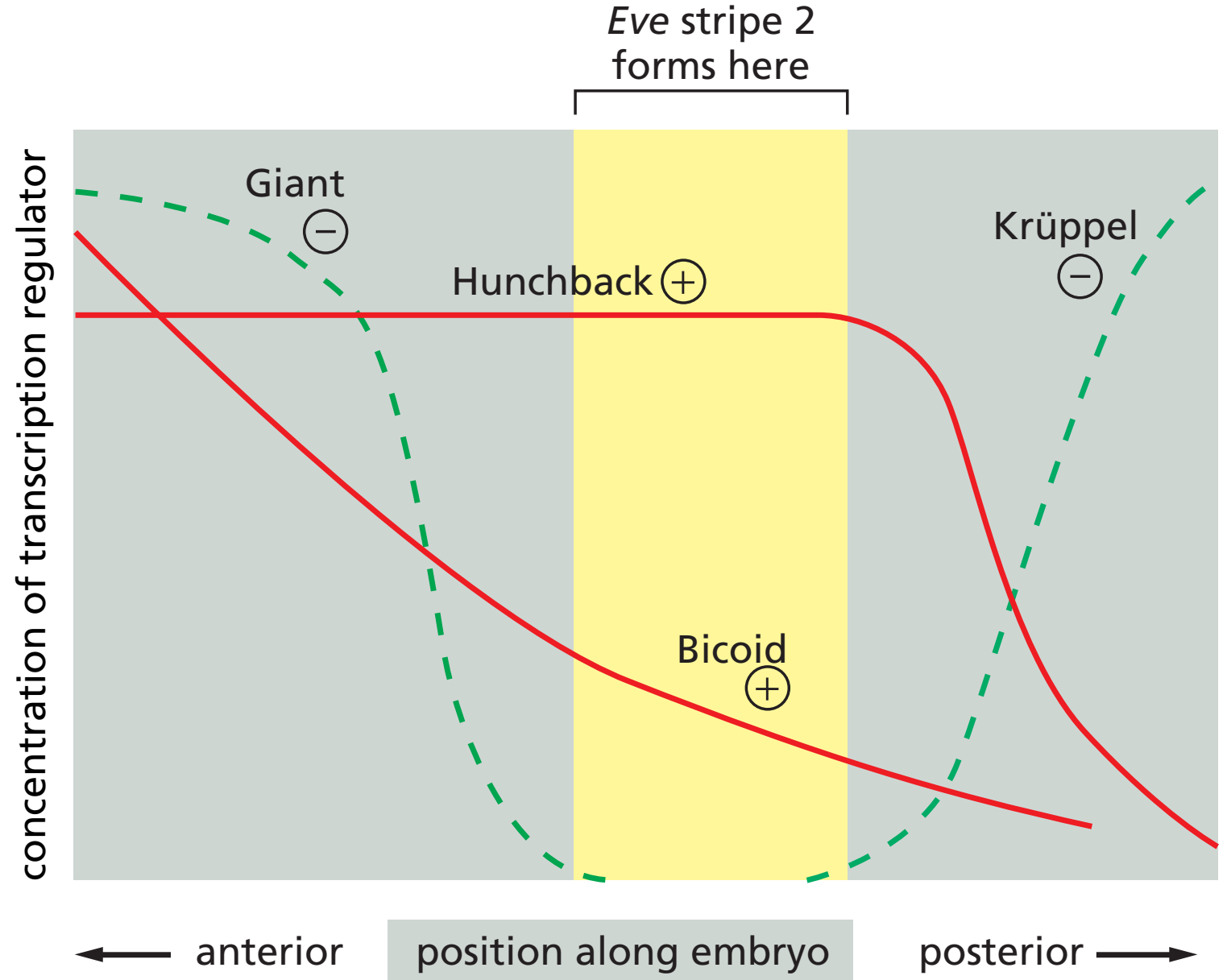
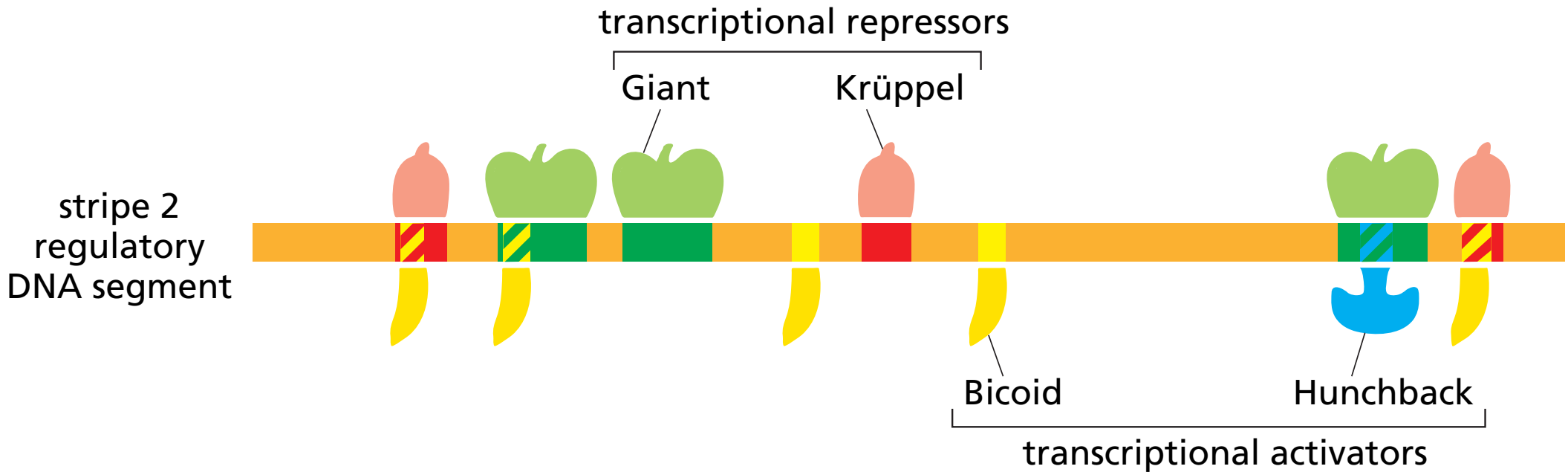
(C)



(D)

# Drosophila Even-skipped gene

- How is the positioning in each stripe controlled?



# IV. Control of gene expression

## 1. Transcriptional control

a. Transcription regulators (activators, repressors)

**b. Understanding different control systems**

c. Combinatorial gene control and cell types



a. Tryptophan repressor (bacteria)

b. The lac operon (bacteria)

c. The drosophila even-skipped gene

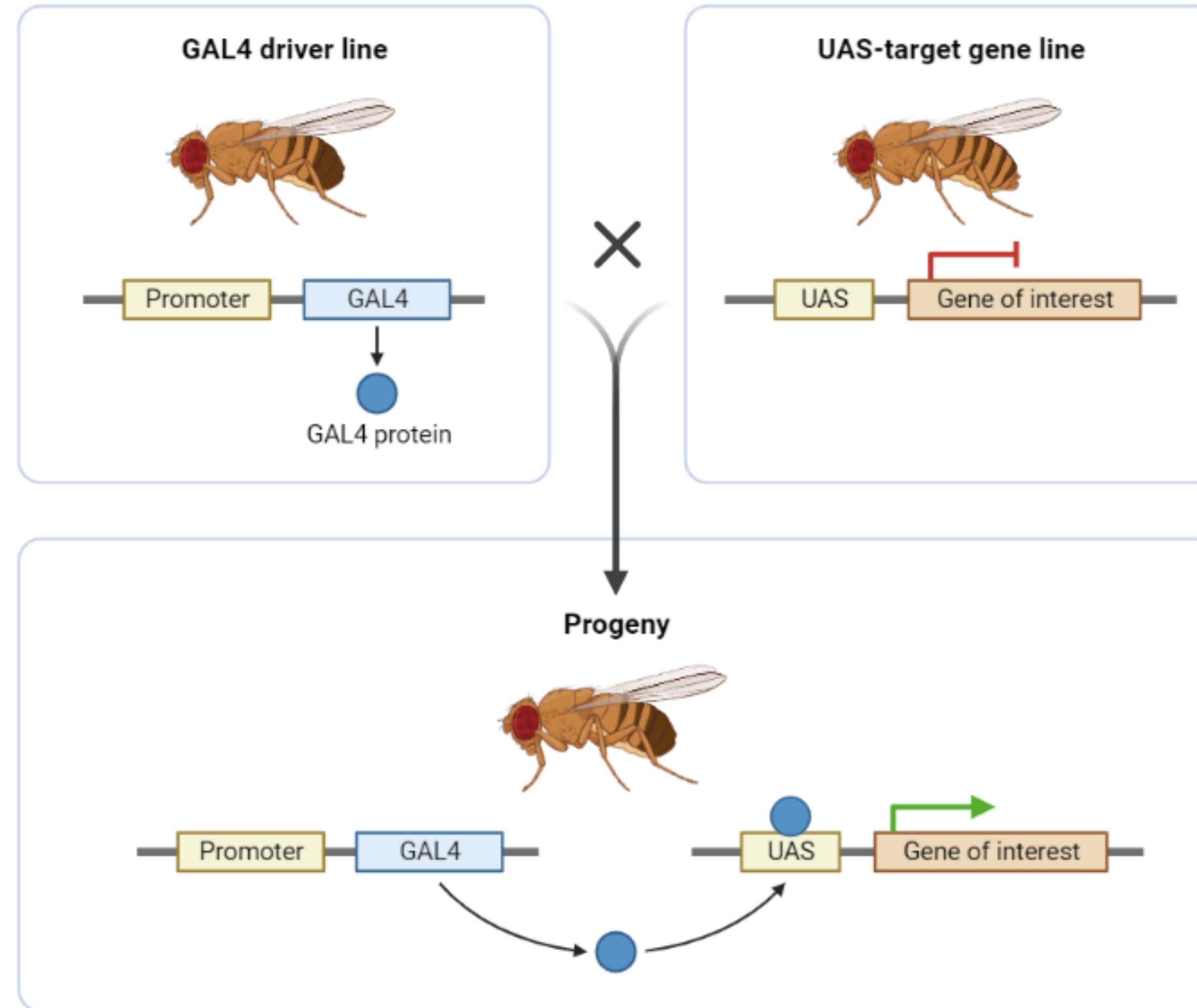
**d. The Gal4-UAS from yeast**

e. The Arabinose operon (bacteria)

# The Gal4-UAS

- The Gal4-UAS system is a two-part genetic tool that allows scientists to control where and when a gene is expressed (used mostly in *Drosophila*, but also in other model organisms)
- It is derived from yeast (*Saccharomyces cerevisiae*), where it regulates the galactose metabolism and it consists of
  - **Gal4** – a transcriptional activator protein.
  - **UAS** – the “Upstream Activating Sequence,” which Gal4 binds to.
- The Gal4 “Driver” Line
  - The Gal4 gene is placed under the control of a specific promoter (a DNA sequence that determines where and when a gene is active).
  - This promoter can be tissue-specific (e.g., in neurons, muscles, or eyes) or developmental stage-specific
  - Result: Gal4 protein is produced only in cells where the promoter is active.
- The UAS “Responder” Line
  - This line carries a gene of interest (often called a “transgene”) downstream of the UAS sequence.
  - Example: UAS-GFP (green fluorescent protein), UAS-RNAi, UAS-GeneX, etc.
  - The UAS itself does nothing unless Gal4 is present.
- When You Cross Them
  - In the offspring:
    - Gal4 is expressed wherever the promoter dictates.
    - Gal4 binds to the UAS sequences.
    - This activates transcription of the gene downstream of the UAS.
    - Result: Your gene of interest is expressed only in the cells where the driver’s promoter is active.

# The Gal4-AUS



# The Gal4-AUS

<b>Purpose</b>	<b>Example</b>
Targeted gene overexpression	Overexpress a growth factor in the developing eye to study tissue development.
RNA interference (RNAi)	Knock down a specific gene in neurons to study its role in behavior.
Reporter expression	Express GFP or lacZ in certain cells to visualize their shape or lineage.
Conditional or temporal control	Combine with temperature-sensitive GAL80 (GAL80 <sup>ts</sup> ) to control when expression happens.

# The Gal4-AUS

<b>Organism</b>	<b>Common Use of GAL4–UAS System</b>
Yeast	Original system; gene regulation studies
Drosophila	Developmental genetics, neurobiology, gene function
Zebrafish	Developmental and neural studies
Mouse	Tissue-specific gene expression
Plants	Gene function and signaling studies
Others (worms, frogs, etc.)	Specialized applications or experimental use

# IV. Control of gene expression

## 1. Transcriptional control

a. Transcription regulators (activators, repressors)

**b. Understanding different control systems**

c. Combinatorial gene control and cell types



a. Tryptophan repressor (bacteria)

b. The lac operon (bacteria)

c. The drosophila even-skipped gene

d. The Gal4-UAS from yeast

**e. The Arabinose operon (bacteria)**

# The Arabinose operon

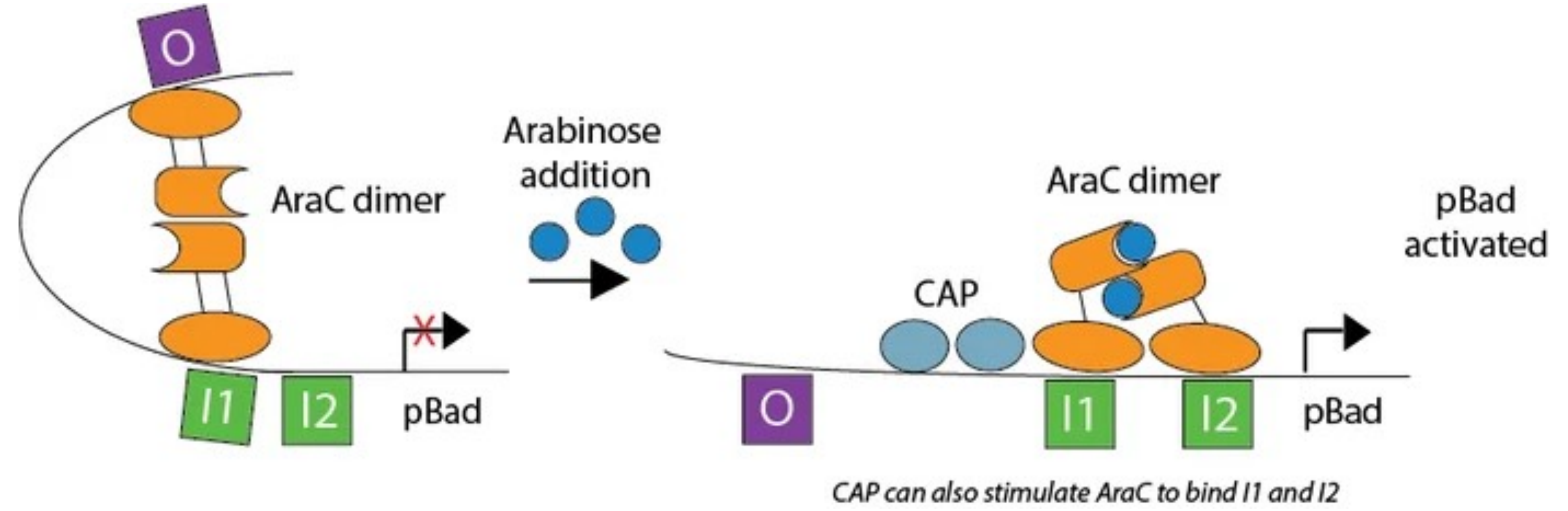
- The **arabinose (ara) operon** is a set of genes in *Escherichia coli (E. coli)* that allows the bacterium to **metabolize the sugar L-arabinose** when it's available in the environment.
- The arabinose operon acts as a **genetic switch** that turns **ON** when arabinose is present and **OFF** when it's absent.

Component	Type	Function
<b>araC</b>	Regulatory gene	Produces the AraC protein, which controls transcription of the operon.
<b>PBAD</b>	Promoter	Where RNA polymerase binds to transcribe the structural genes.
<b>O2, O1, I1, I2</b>	Operator sites	DNA sequences where AraC binds to regulate the operon.
<b>araB, araA, araD</b>	Structural genes	Encode enzymes that convert arabinose into D-xylulose-5-phosphate, an intermediate in the pentose phosphate pathway.
<b>CAP site</b>	Binding site for the CAP-cAMP complex	Enhances transcription when glucose is low.

# The Arabinose operon

- In the absence of arabinose — operon OFF
  - **AraC protein** forms a **DNA loop** by binding to two sites: **O<sub>2</sub>** and **I<sub>1</sub>**.
  - This loop blocks RNA polymerase from binding to the PBAD promoter.
  - Result: No transcription of *araB*, *araA*, or *araD*.
  - The cell doesn't waste energy producing enzymes it doesn't need.
- In the presence of arabinose — operon ON
  - Arabinose binds to the **AraC protein**, changing its shape (a conformational change).
  - The AraC–arabinose complex binds to **I<sub>1</sub>** and **I<sub>2</sub>** instead of looping the DNA.
  - This opens up the promoter region and **allows RNA polymerase** to access **PBAD**.
  - The operon is **activated**, and transcription of **araB**, **araA**, and **araD** begins
- Glucose effect (catabolite repression)
  - Even if arabinose is present, if **glucose is also available**, transcription stays **low**.
  - That's because the **CAP–cAMP complex** is needed to fully activate **PBAD**, and cAMP levels are low when glucose is high.
  - So the operon only runs strongly when arabinose is present and glucose is absent — a form of metabolic efficiency.

# The Arabinose operon



# Plan

- Different cell types have the same DNA
- Transcriptional control
  - Transcriptional regulators
    - Activators
    - Repressors
  - Understanding other regulatory systems
- **Combinatorial gene control and cell types**

# IV. Control of gene expression

## 1. Transcriptional control

a. Transcription regulators (activators, repressors)

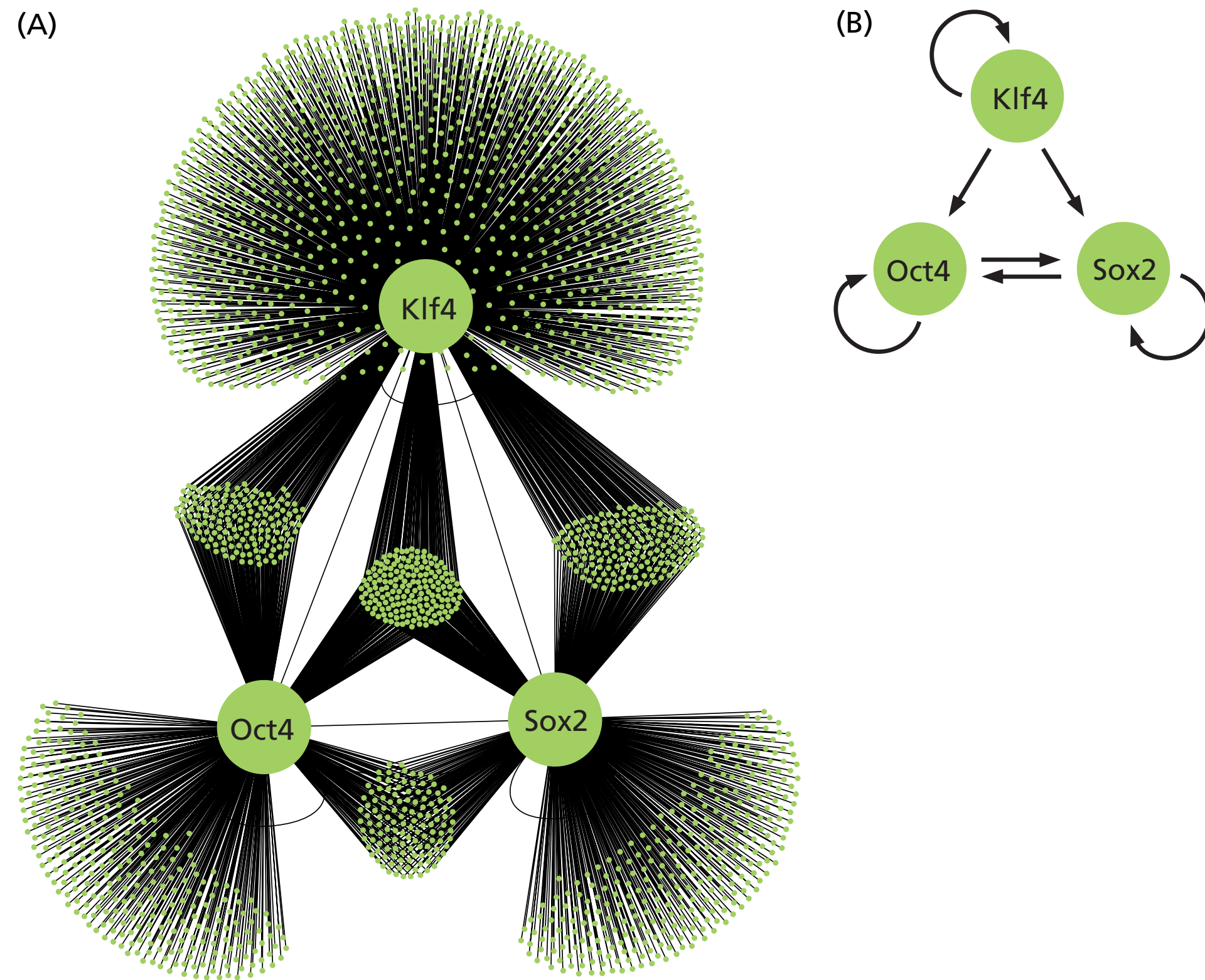
b. Understanding different control systems

**c. Combinatorial gene control and cell types**

# Master transcription regulators

- Different cell types express different proteins
- Controlled by **master transcription regulators**
- Master transcription regulators (also called master transcription factors or master regulators) are special **transcription factors** that control **large gene networks** and play decisive roles in **determining cell identity, developmental fate, and lineage specification**
- A **master transcription regulator** is a transcription factor that:
  - **Binds to DNA** at specific regulatory regions (enhancers/promoters) of many genes.
  - **Controls other transcription factors** — often activating or repressing entire regulatory cascades.
  - **Drives cell fate decisions** — turning on the gene expression program characteristic of a specific cell type or developmental pathway.

# Master transcription regulators



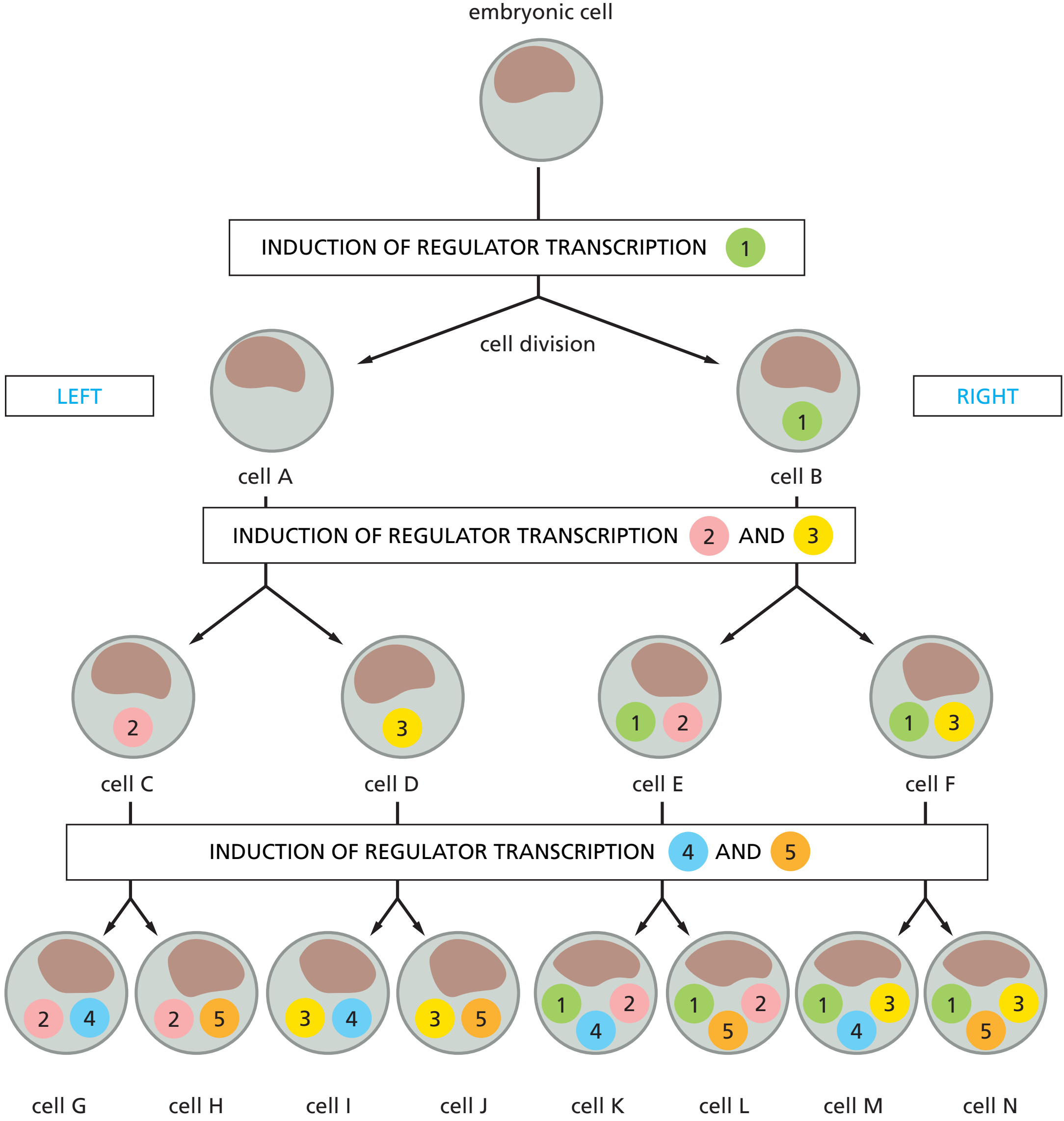
## Mechanistic Insights

- They often bind to “super-enhancers”, large clusters of enhancers that drive expression of genes defining cell identity.
- They work in combinatorial networks — several master regulators together specify a cell type (e.g., OCT4–SOX2–NANOG in pluripotent cells).
- Epigenetic remodeling is a key consequence — they recruit chromatin modifiers to open or close specific genomic regions.

Master regulators are critical in **cellular reprogramming** — for example:

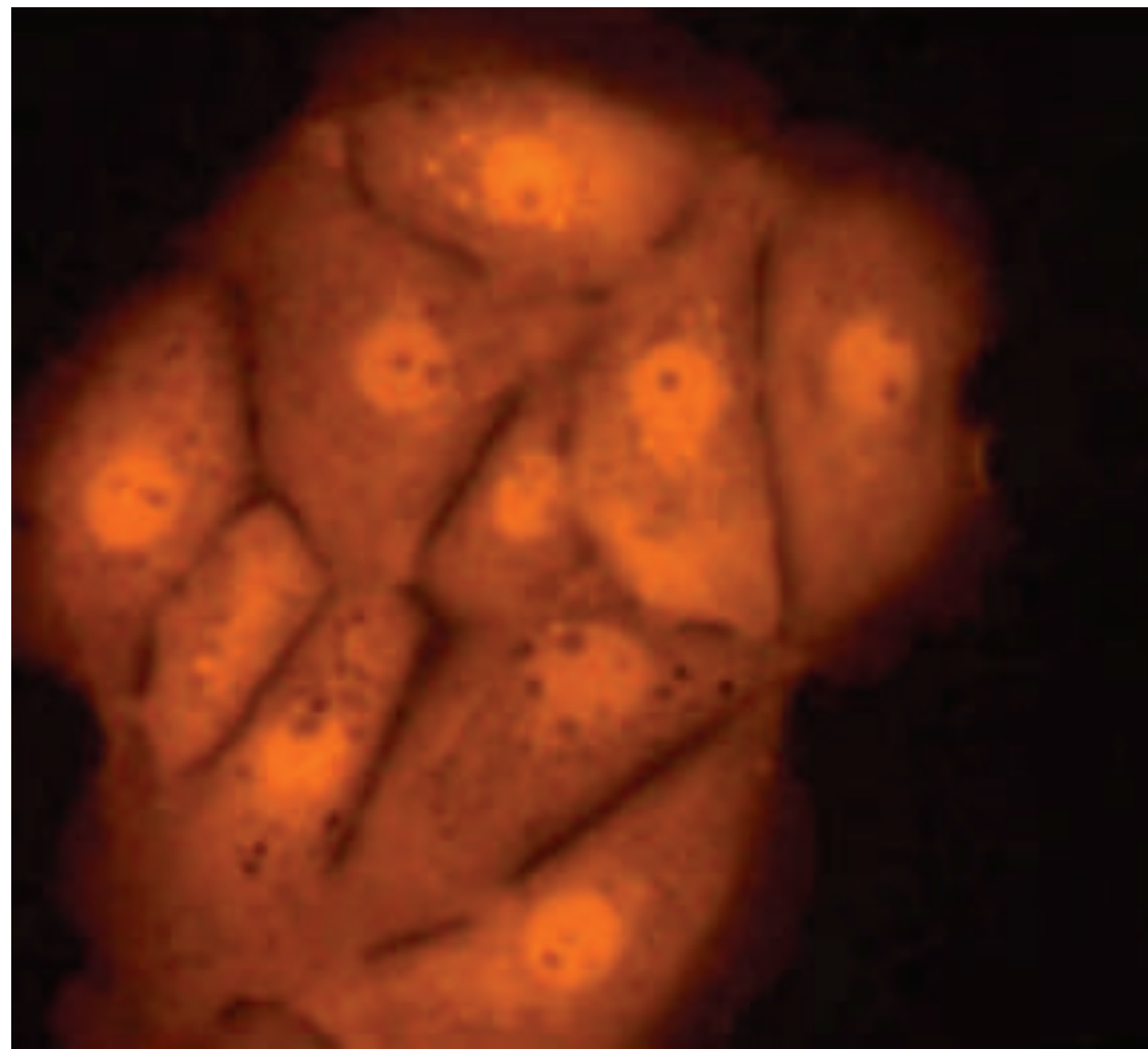
- The Yamanaka factors (OCT4, SOX2, KLF4, c-MYC) can reprogram somatic cells into induced pluripotent stem cells (iPSCs).
- Expression of MyoD alone can convert fibroblasts into myoblasts.

# Combinatorial gene control creates cell types



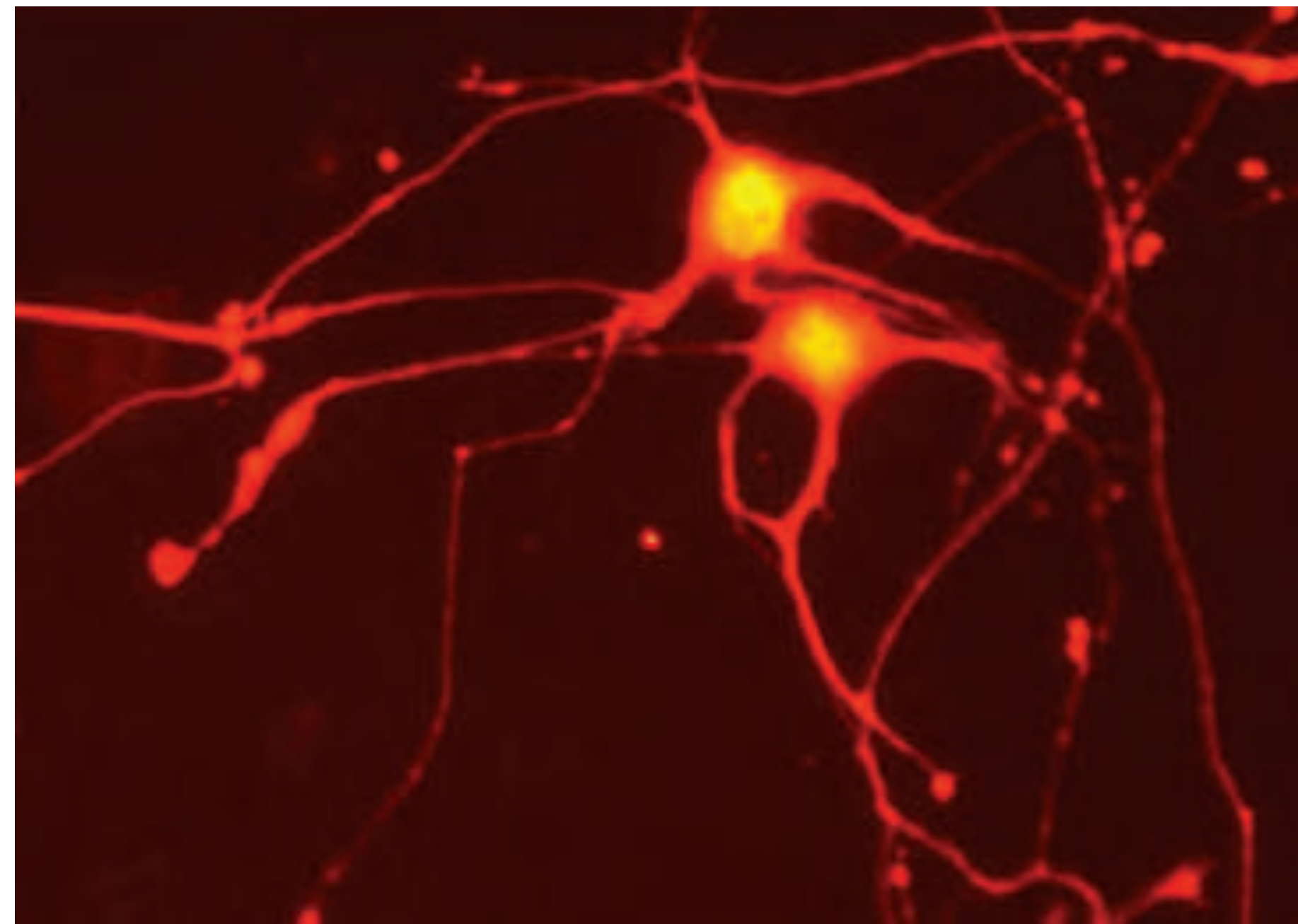
# Combinatorial gene control creates cell types

- Artificial expression of 3 neuron-specific transcription regulators in liver cells converts them in nerve cells



(A)

50 μm



(B)

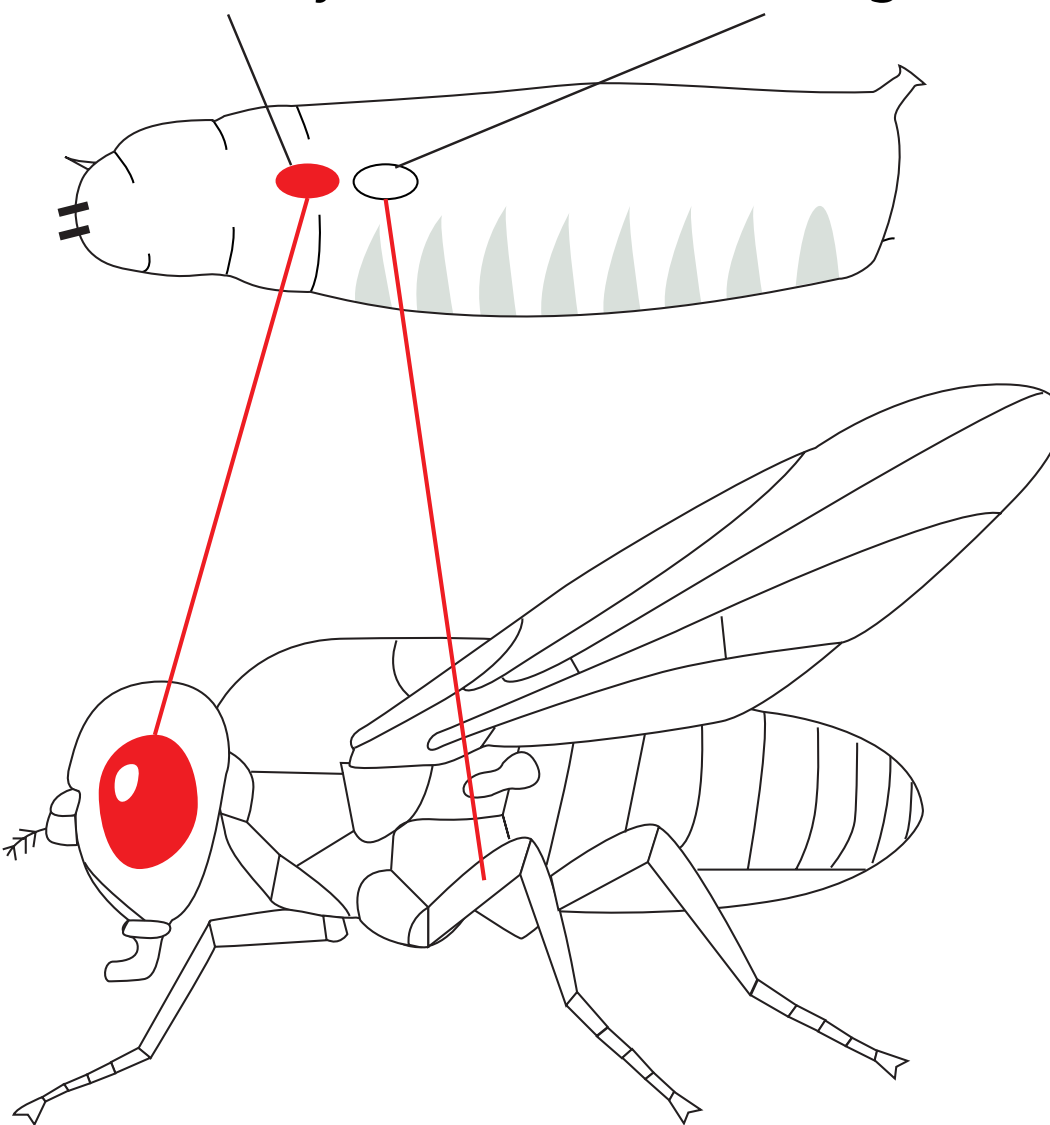
50 μm

# Combinatorial gene control creates cell types

- Artificial expression of 1 transcription regulator from the eye region to the leg region

group of cells that give rise to an adult eye

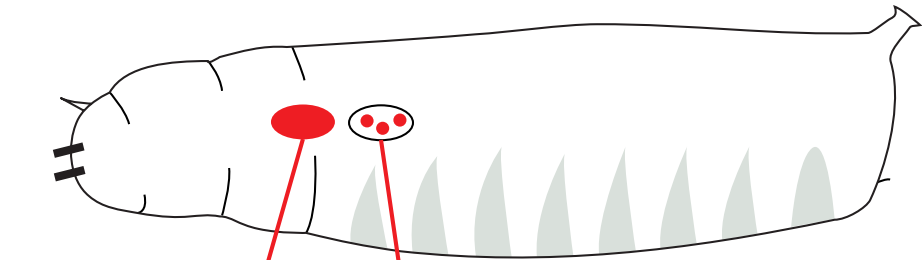
group of cells that give rise to an adult leg



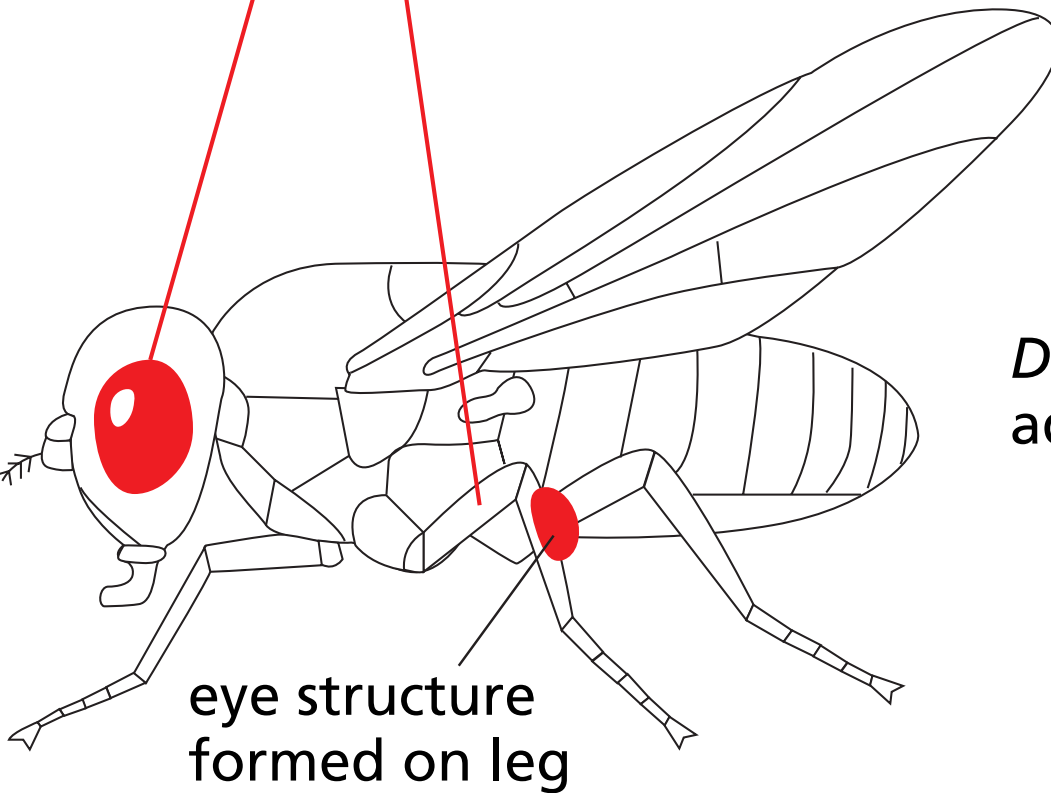
normal fly

(A)

(red shows cells expressing *Eyeless* gene)

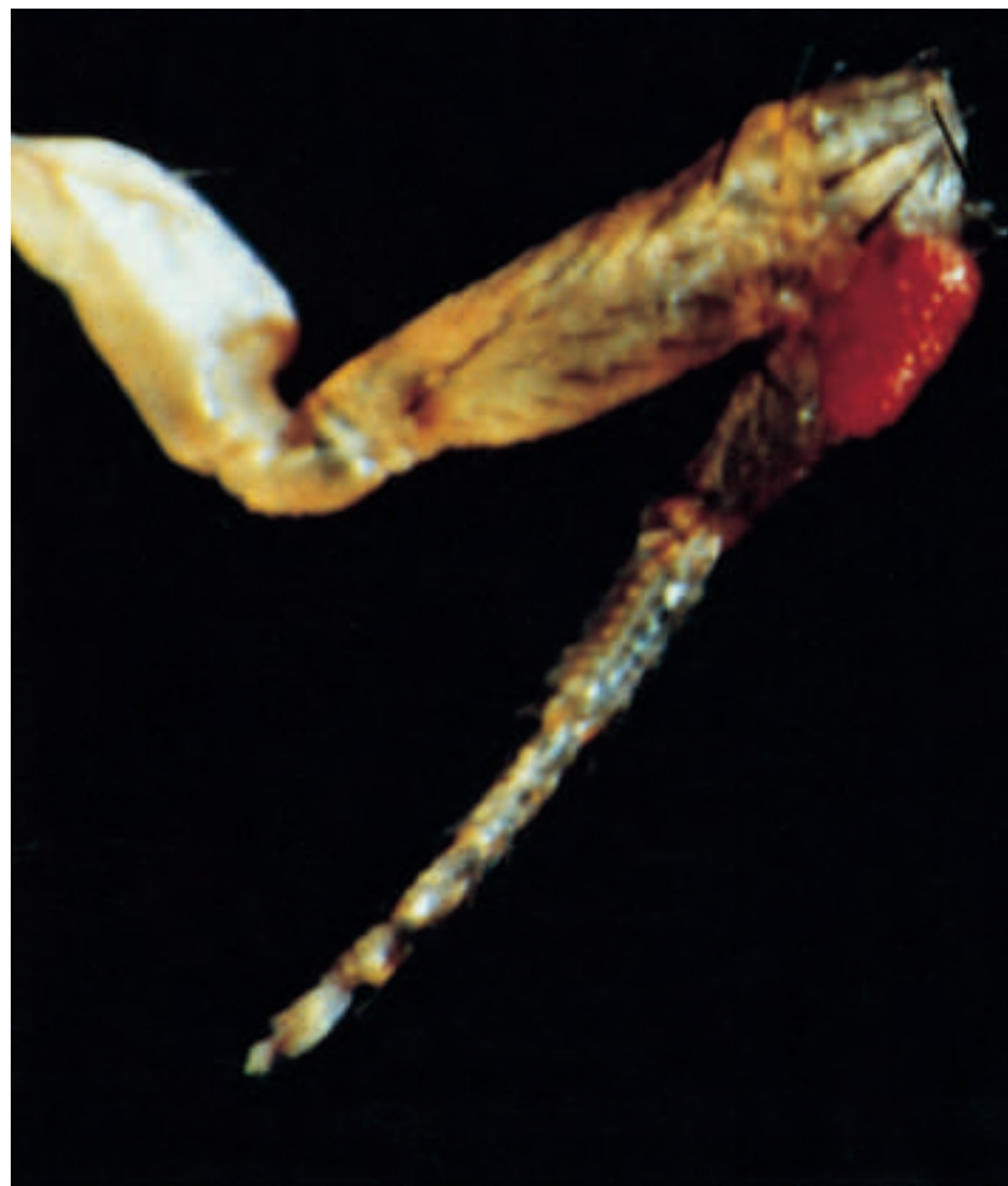


*Drosophila* larva



*Drosophila* adult

fly with *Eyeless* gene artificially expressed in leg precursor cells

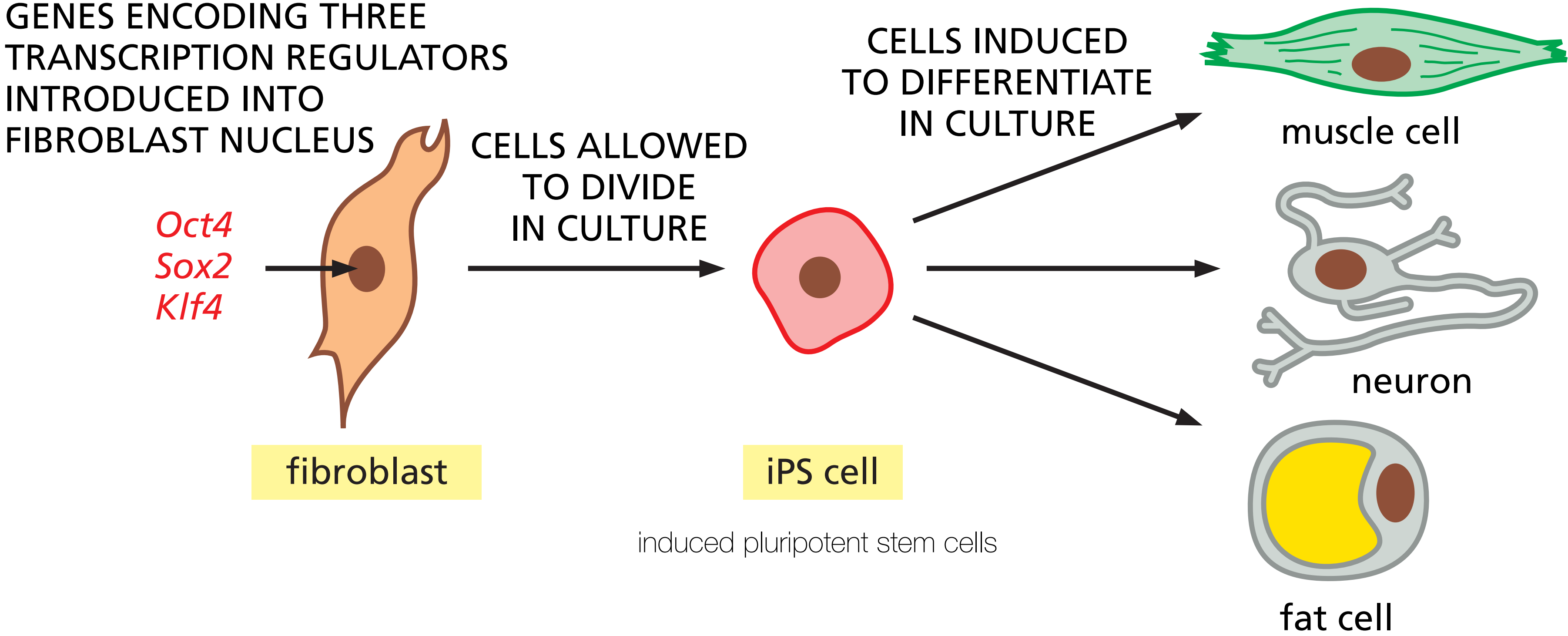


(B)



# Pluripotent stem cells

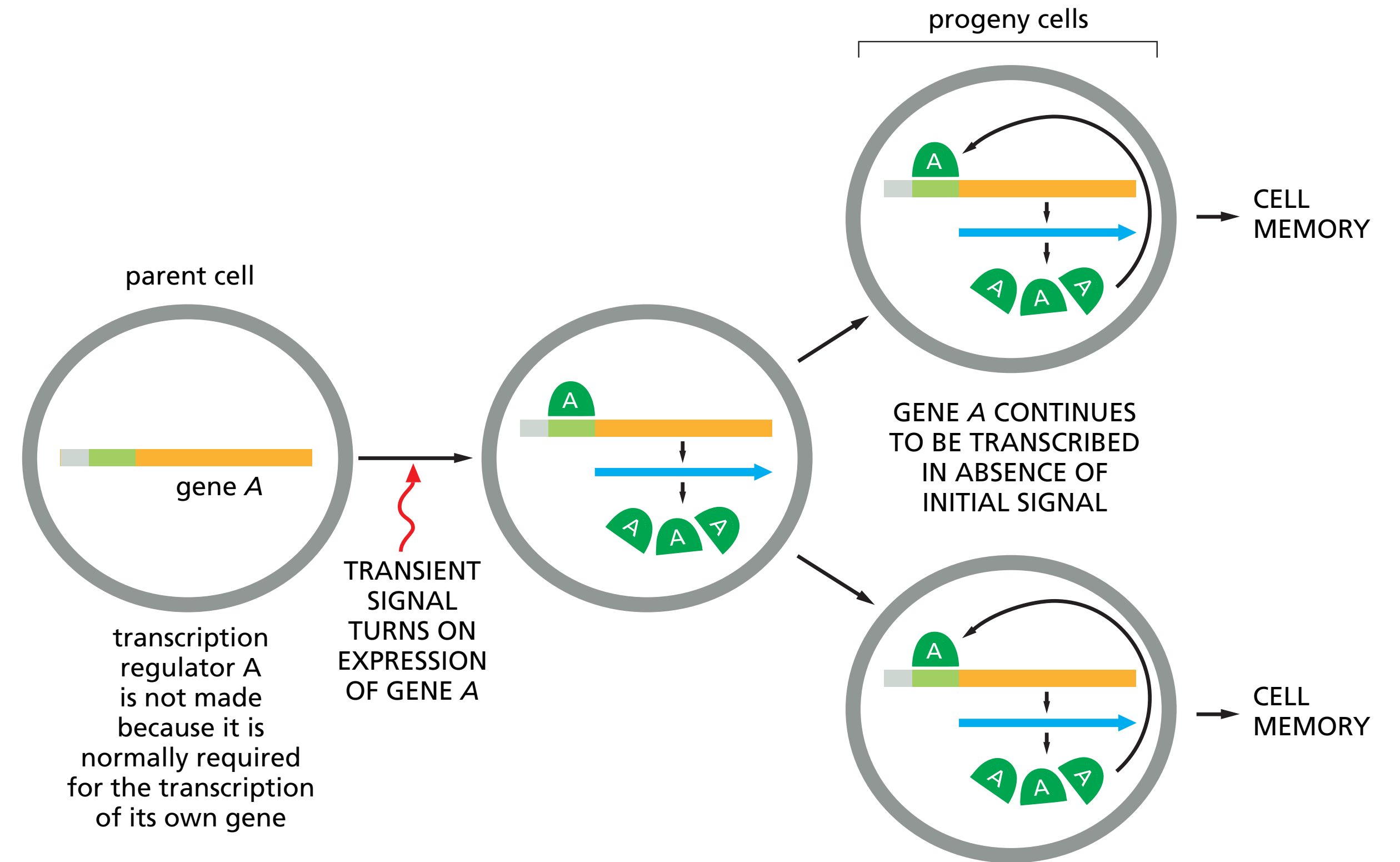
- Manipulation of **transcription regulators** to de-differentiate cells (see later)



# Maintaining cell types

Differentiated cells maintain **their identity**

- Progeny will remain the **same cell type - cell memory**
- Some are **terminally differentiated** (no further division) like neurons or skeletal muscle cells
- **Positive feedback loop** so a master transcription regulator activates transcription of its own gene

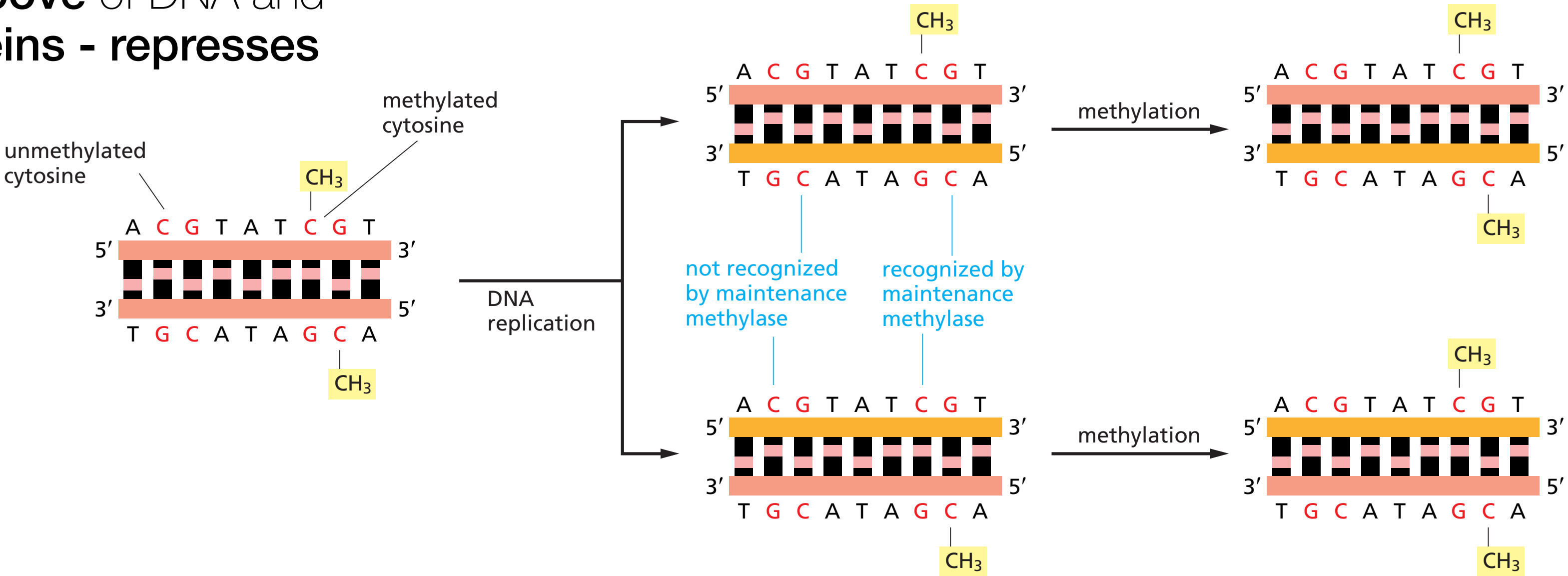
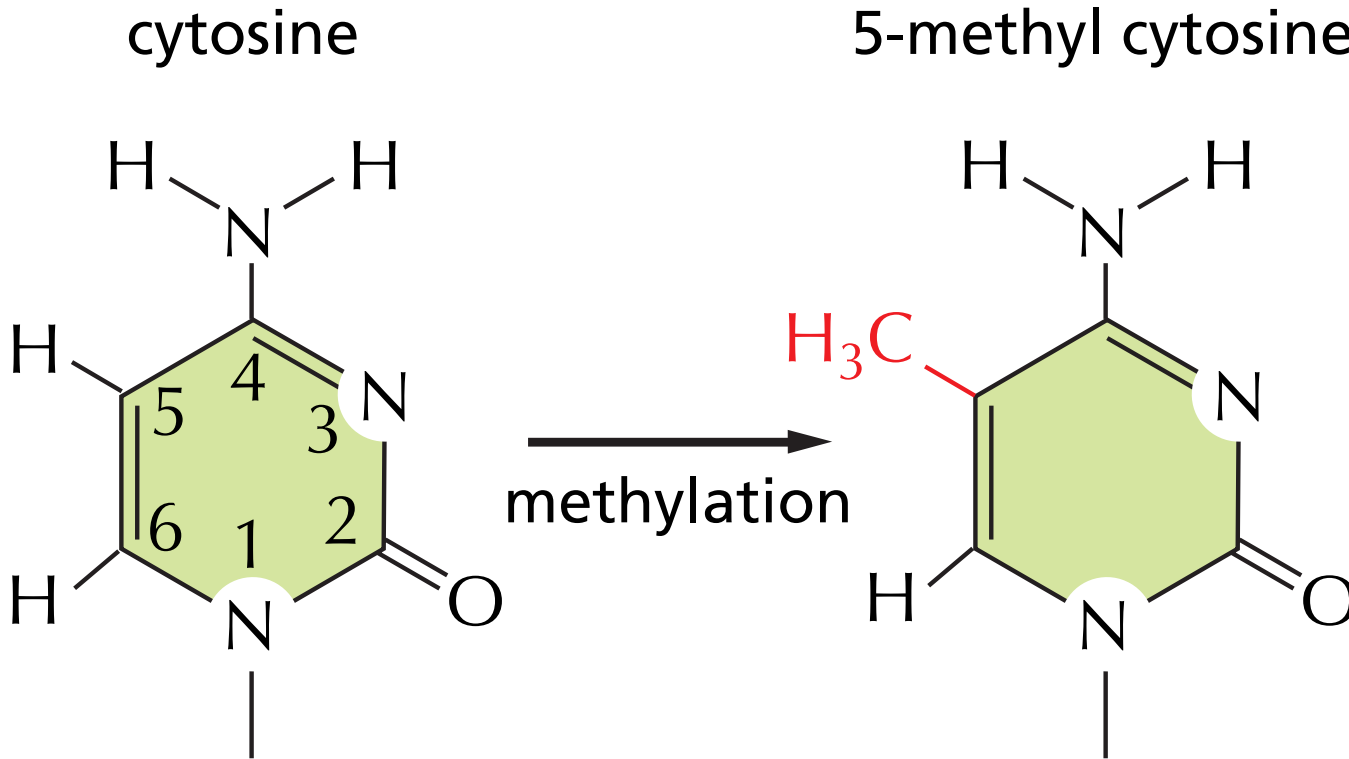


Concept: A **positive feedback loop** is a process in which the output of a system amplifies or reinforces the initial input, leading to an **increase in the overall effect**.

# Maintaining cell types - cell memory

DNA methylation can be **inherited** when cells divide

- **DNA methylation** occurs on **cytosine** (largely in sequence CG)
- **Maintenance methyl transferases** act on these newly-made CG sequences (paired with methylated CG sequences)
- Methyl group lies in the **major groove** of DNA and interfere with the **binding of proteins - represses transcription**

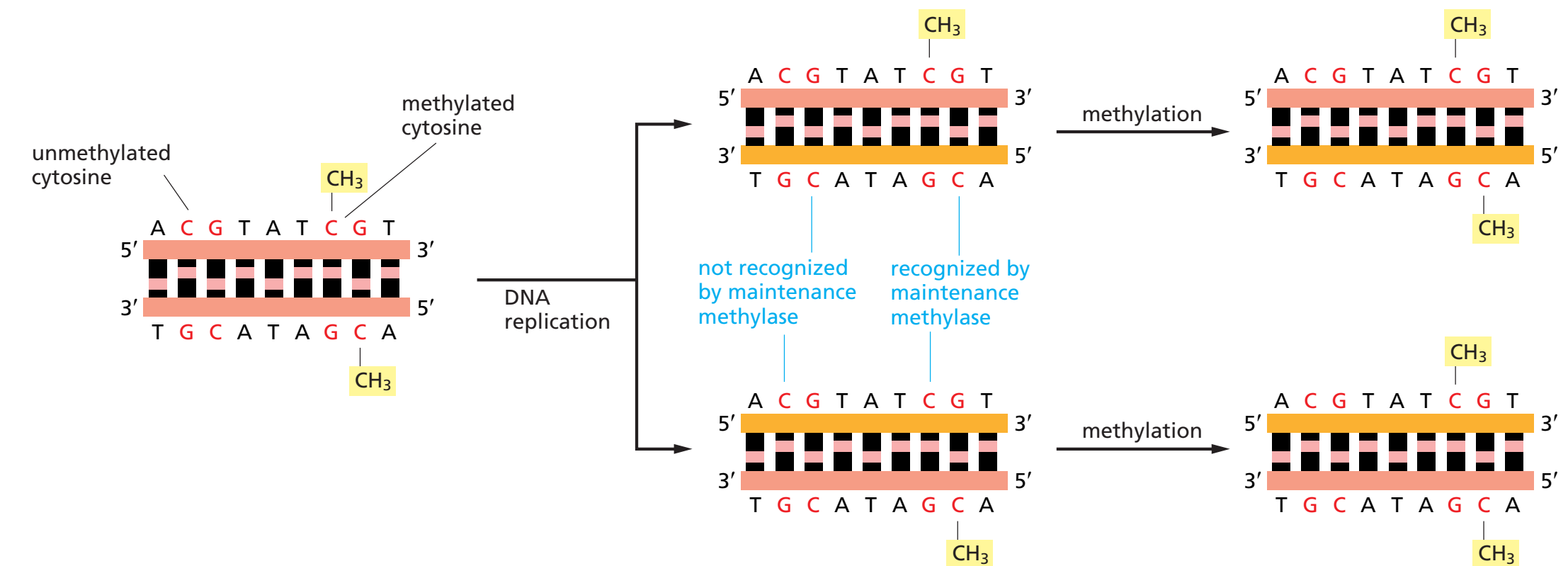
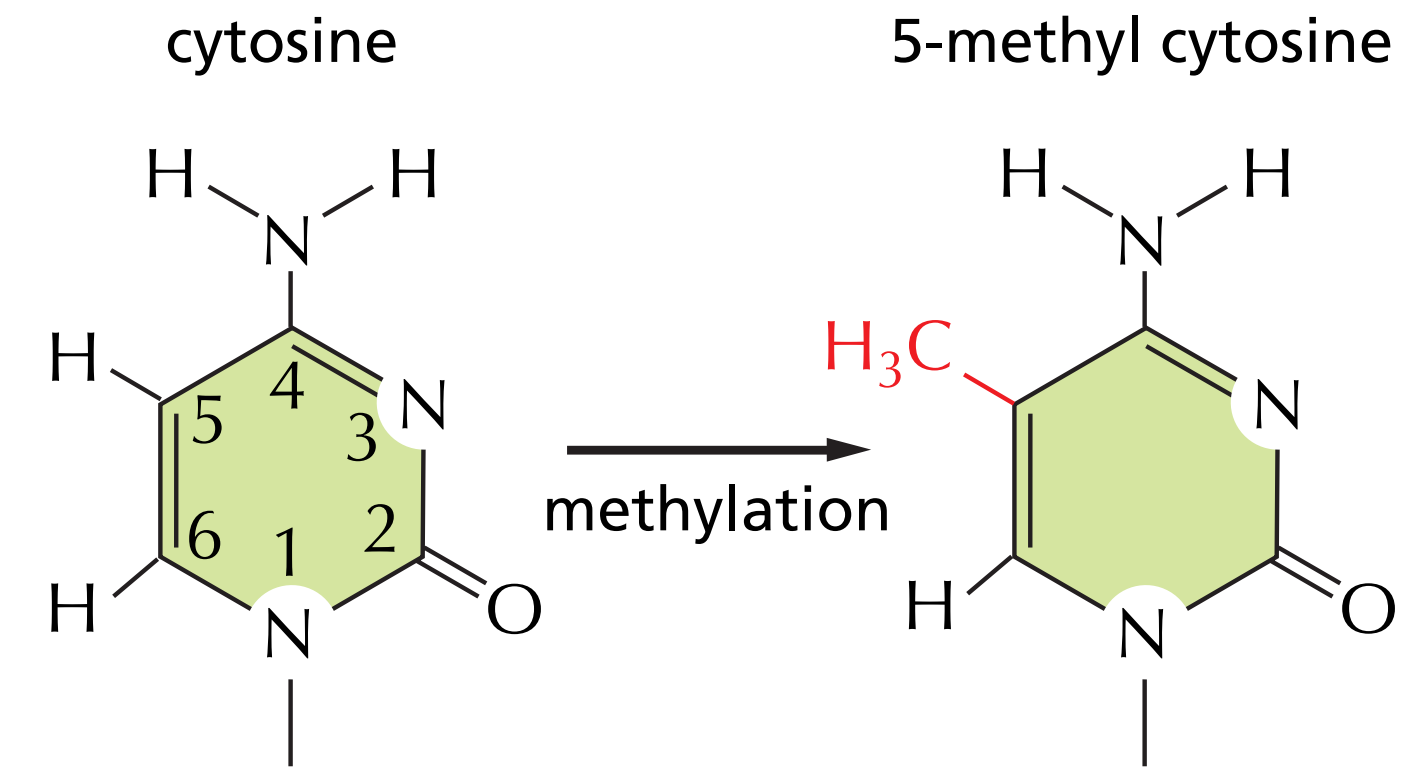


# Maintaining cell types - cell memory

DNA methylation can be **inherited** when cells divide

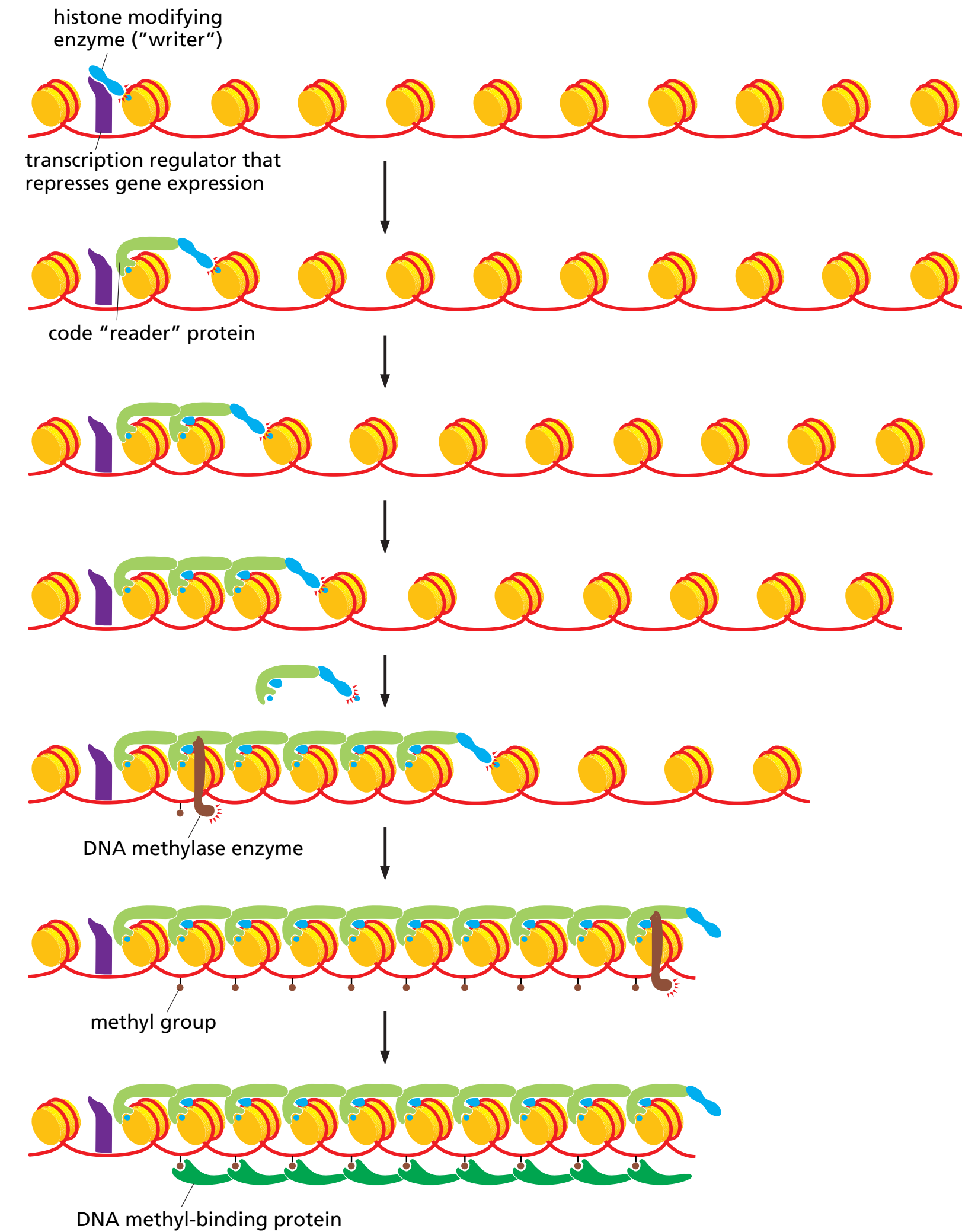
- **DNA methylation** occurs on **cytosine** (largely in sequence CG)
- **Maintenance methyl transferases** act on these newly-made CG sequences (paired with methylated CG sequences)
- Methyl group lies in the **major groove** of DNA and interfere with the **binding of proteins - represses transcription**

**Histone modifications** can also be inherited



# Maintaining cell types - cell memory

- Some proteins bind specifically to **methylated DNA**
- E.g. **histone modifying enzymes** leading to **repressive chromatin state**
- **Synergistic** action of chromatin structure and DNA methylation



**Have a nice day!**