Lecture Series 8

The Eukaryotic Genome and Its Expression

Reading Assignments

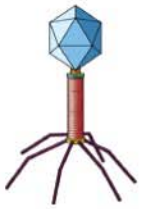
- Read Chapter 8

  Control of Gene Expression
- Skim Chapter 9

  How Genes and Genomes Evolve

A. The Eukaryotic Genome

- Although eukaryotes have more DNA in their genomes than prokaryotes, in some cases there is NO apparent relationship between genome size and organism complexity.



Bacteriophage
10,000 bp
per cell

Yeast
24 million bp

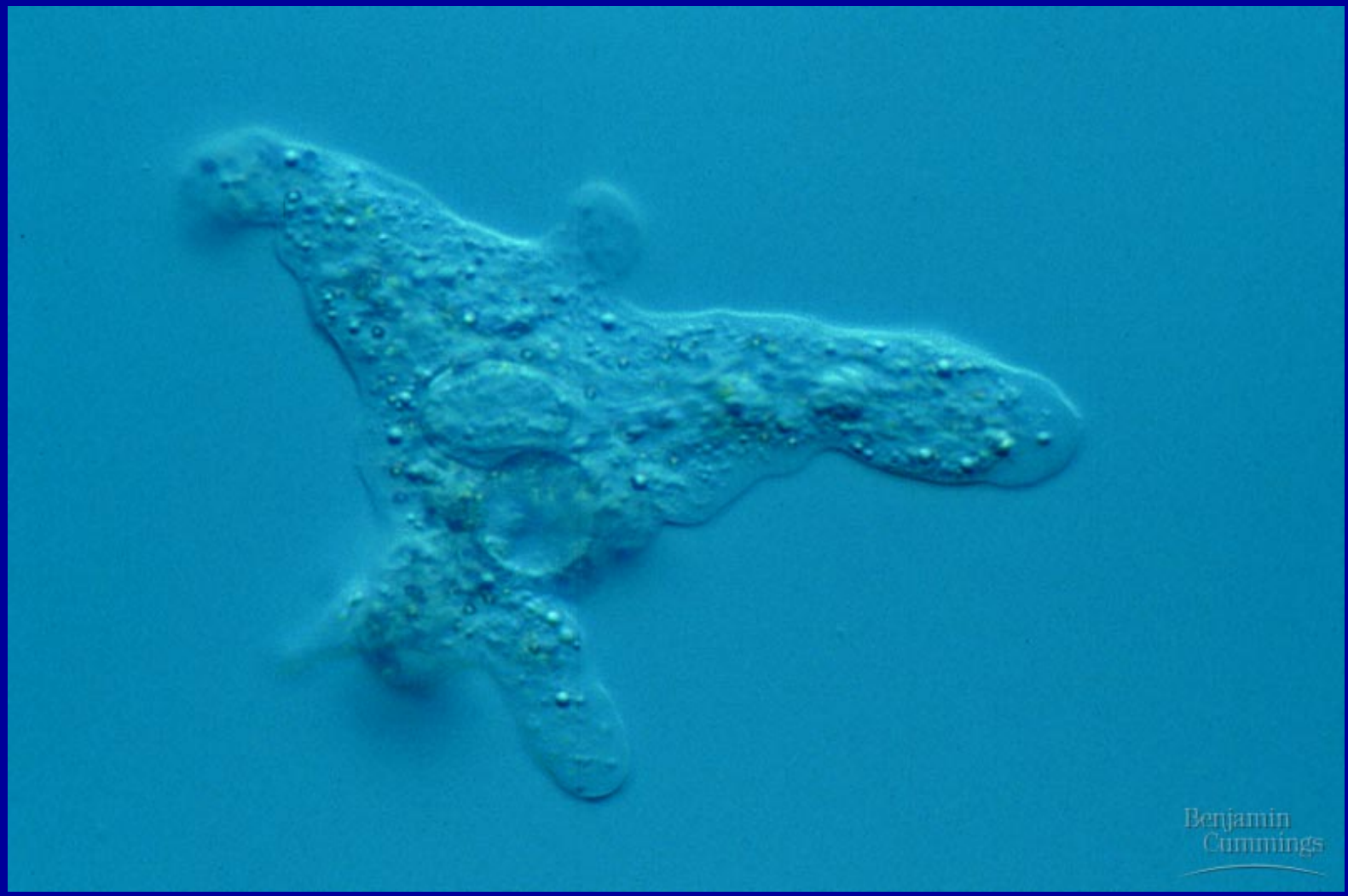
E. coli
4 million bp

Caenorhabditis elegans
160 million bp per cell

Fruit fly
330 million bp

Lily
106 billion bp

Human
6 billion bp
per cell



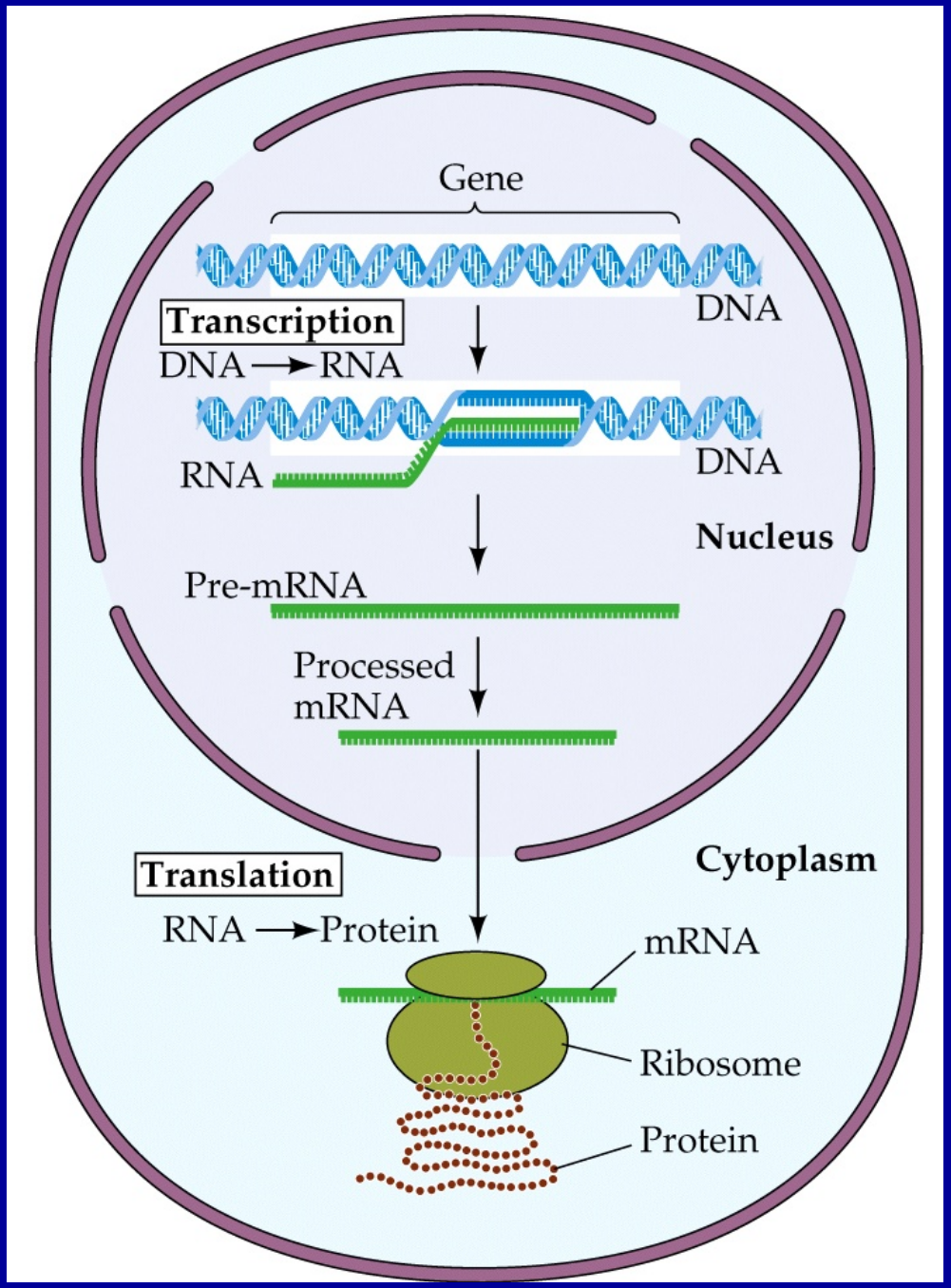
Amoeba dubia is the big winner at **670 Billion** base pairs per cell and an uncertain phylogeny!

14.1 A Comparison of Prokaryotic and Eukaryotic Genes and Genomes

CHARACTERISTIC	PROKARYOTES	EUKARYOTES
Genome size (base pairs)	10^4–10^7	10^8–10^{11}
Repeated sequences	Few	Many
Noncoding DNA within coding sequences	Rare	Common
Transcription and translation separated in cell	No	Yes
DNA segregated within a nucleus	No	Yes
DNA bound to proteins	Some	Extensive
Promoter	Yes	Yes
Enhancer/silencer	Rare	Common
Capping and tailing of mRNA	No	Yes
RNA splicing required	Rare	Common
Number of chromosomes in genome	One	Many

A. The Eukaryotic Genome

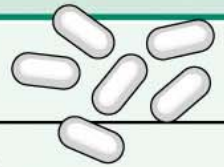
- Unlike prokaryotic DNA, eukaryotic DNA is separated from the cytoplasm by being contained within a nucleus.
- The initial mRNA transcript of the DNA gets modified before it is exported to the cytoplasm.

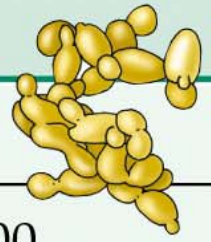

Gene
DNA
Transcription
DNA → RNA
RNA
DNA
Nucleus
Pre-mRNA
Processed
mRNA
Cytoplasm
Translation
RNA → Protein
mRNA
Ribosome
Protein

A. The Eukaryotic Genome

- The genome of the single-celled budding yeast contains genes for the same metabolic machinery as bacteria, as well as genes for protein targeting in the cell.

14.2 Comparison of the Genomes of E. coli and Yeast





	E. COLI	YEAST
Genome length (base pairs)	4,640,000	12,068,000
Number of proteins	4,300	6,200
Proteins with roles in:		
Metabolism	650	650
Energy production/storage	240	175
Membrane transporters	280	250
DNA replication/repair/ recombination	120	175
Transcription	230	400
Translation	180	350
Protein targeting/secretion	35	430
Cell structure	180	250

A. The Eukaryotic Genome

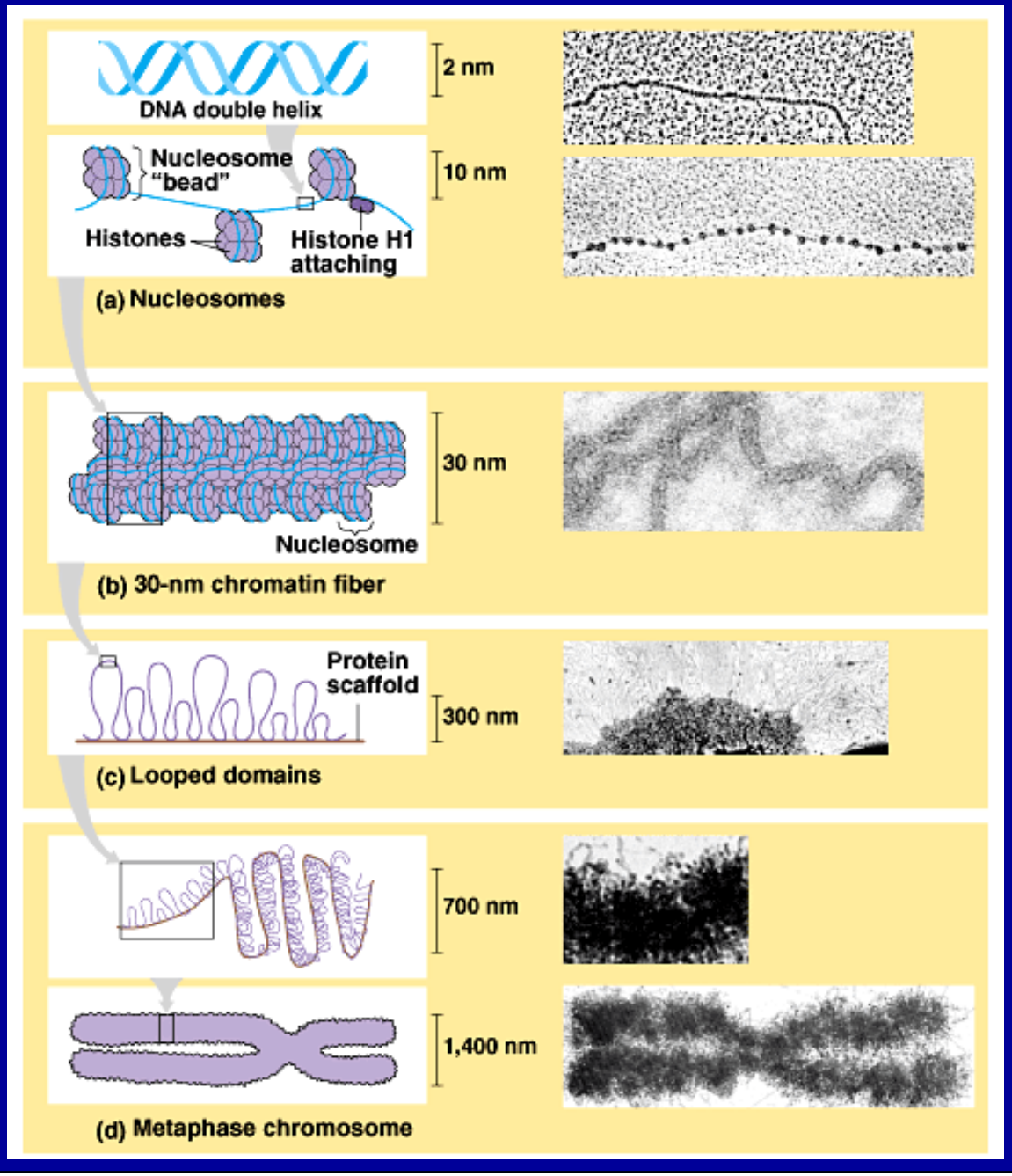
- The genome of the multicellular roundworm *Caenorhabditis elegans* contains genes required for intercellular interactions.
- The genome of the fruit fly has fewer genes than that of the roundworm. Many of its sequences are homologs of sequences on roundworm and mammalian genes.

14.3 C. elegans *Genes Essential to Multicellularity*

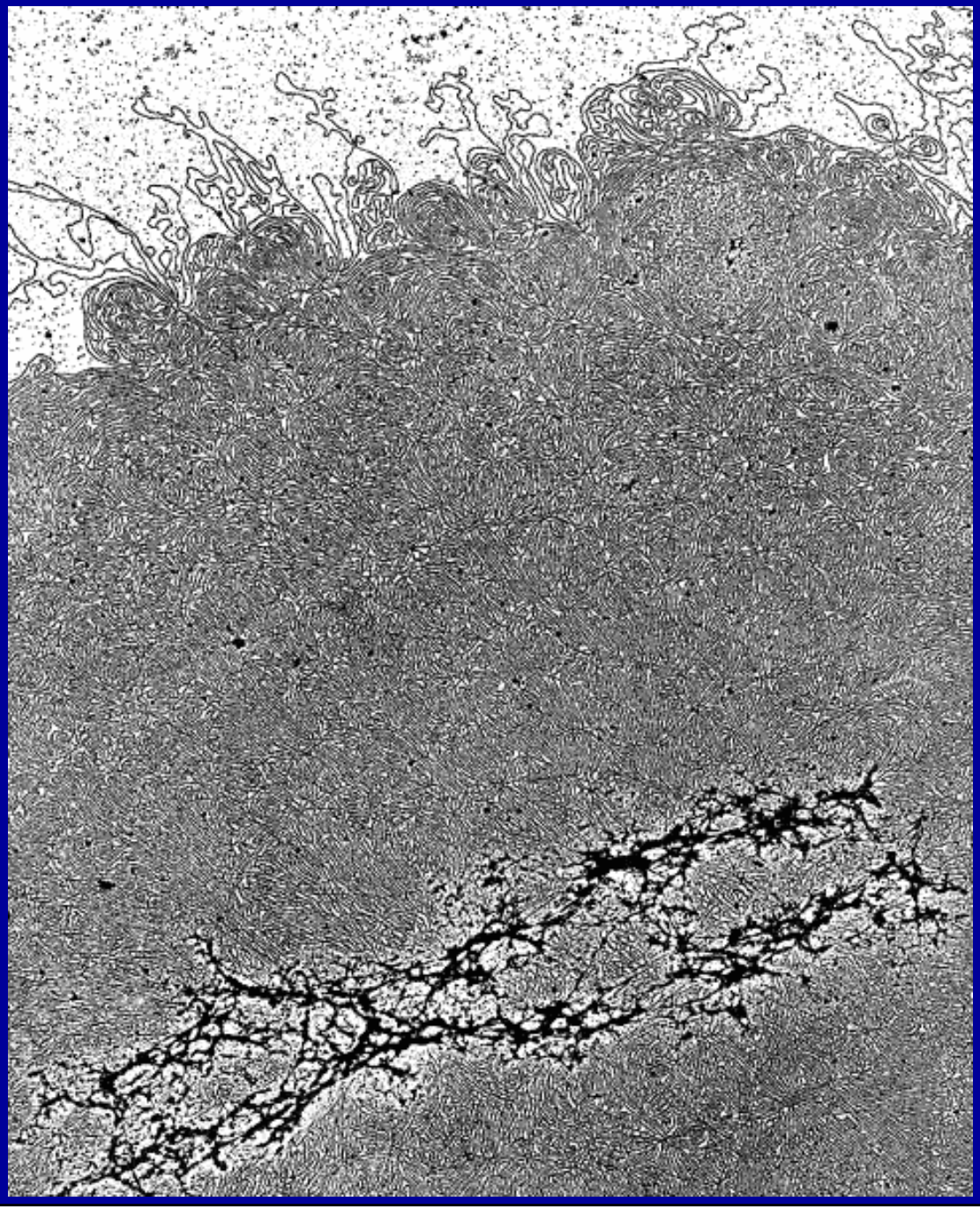
FUNCTION	PROTEIN/DOMAIN	GENES
Transcription control	Zinc finger; homeobox	540
RNA processing	RNA binding domains	100
Nerve impulse transmission	Gated ion channels	80
Tissue formation	Collagens	170
Cell interactions	Extracellular domains; glycotransferases	330
Cell–cell signaling	G protein-linked receptors; protein kinases; protein phosphatases	1,290

Chromatin in a developing salamander ovum

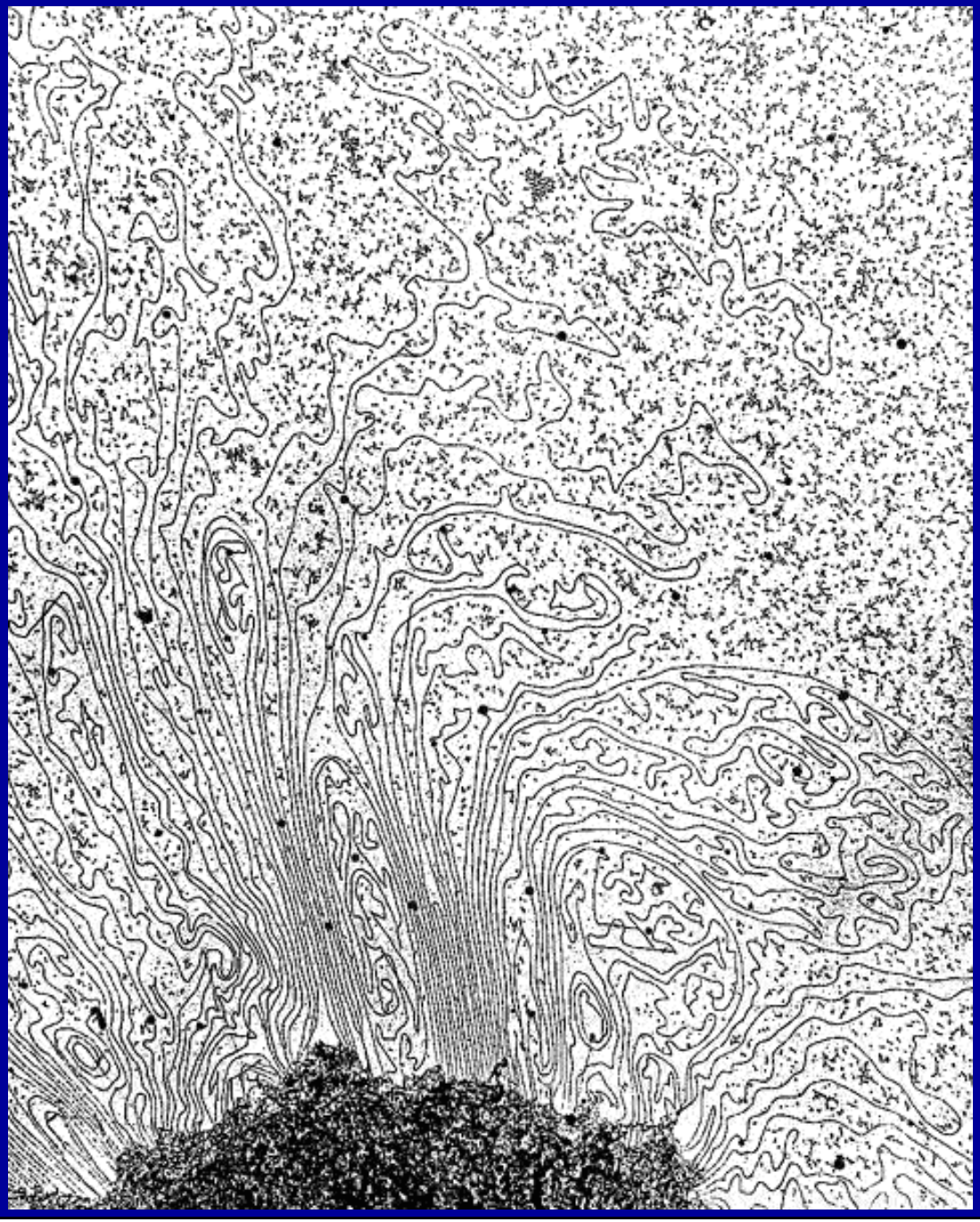
Levels of chromatin packing



Chromatin



Chromatin, detail



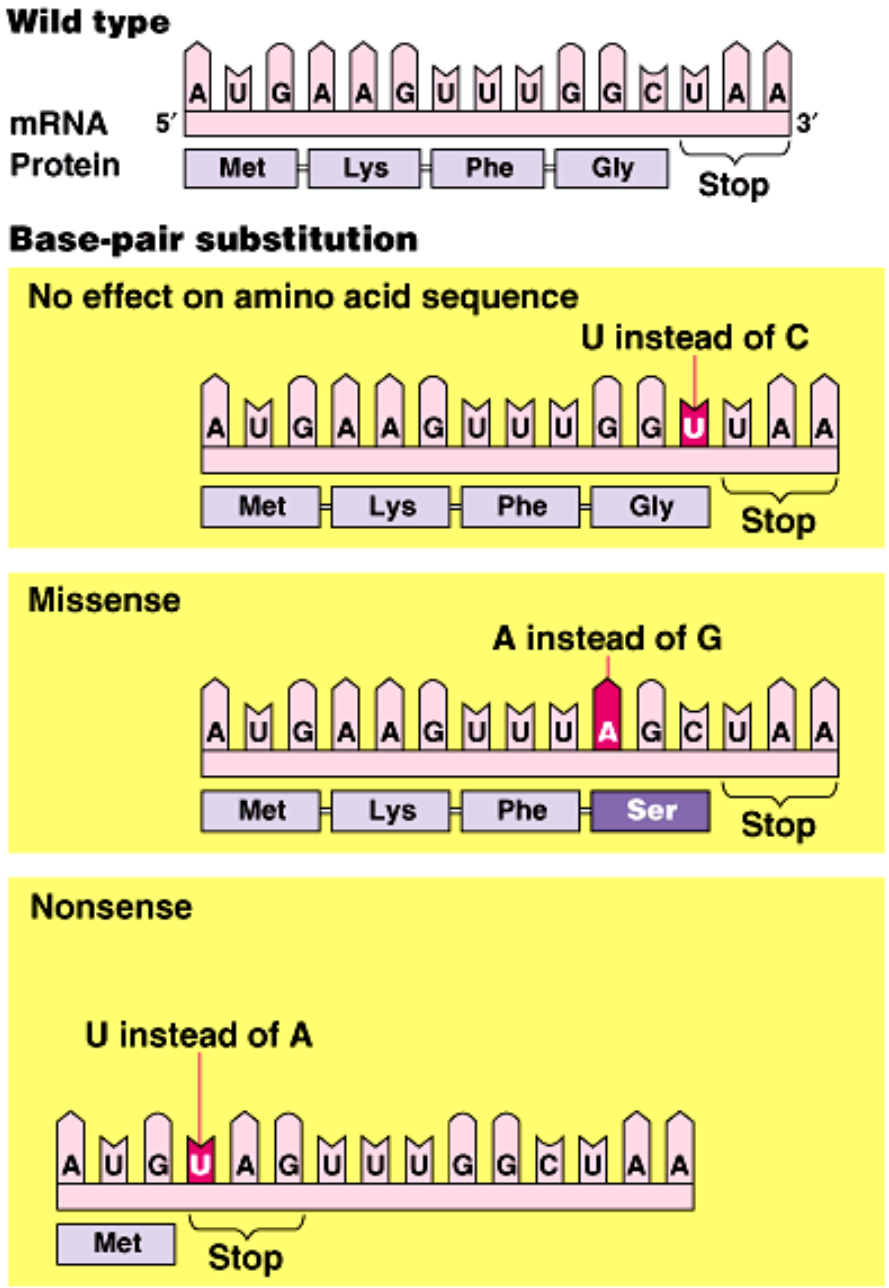
B. Mutations: Heritable Changes in Genes

- Mutations in DNA are often expressed as abnormal proteins. However, the result may not be easily observable phenotypic changes.
- Raw materials for evolution to operate.
- Some mutations appear only under certain conditions, such as exposure to a certain environmental agent or condition.

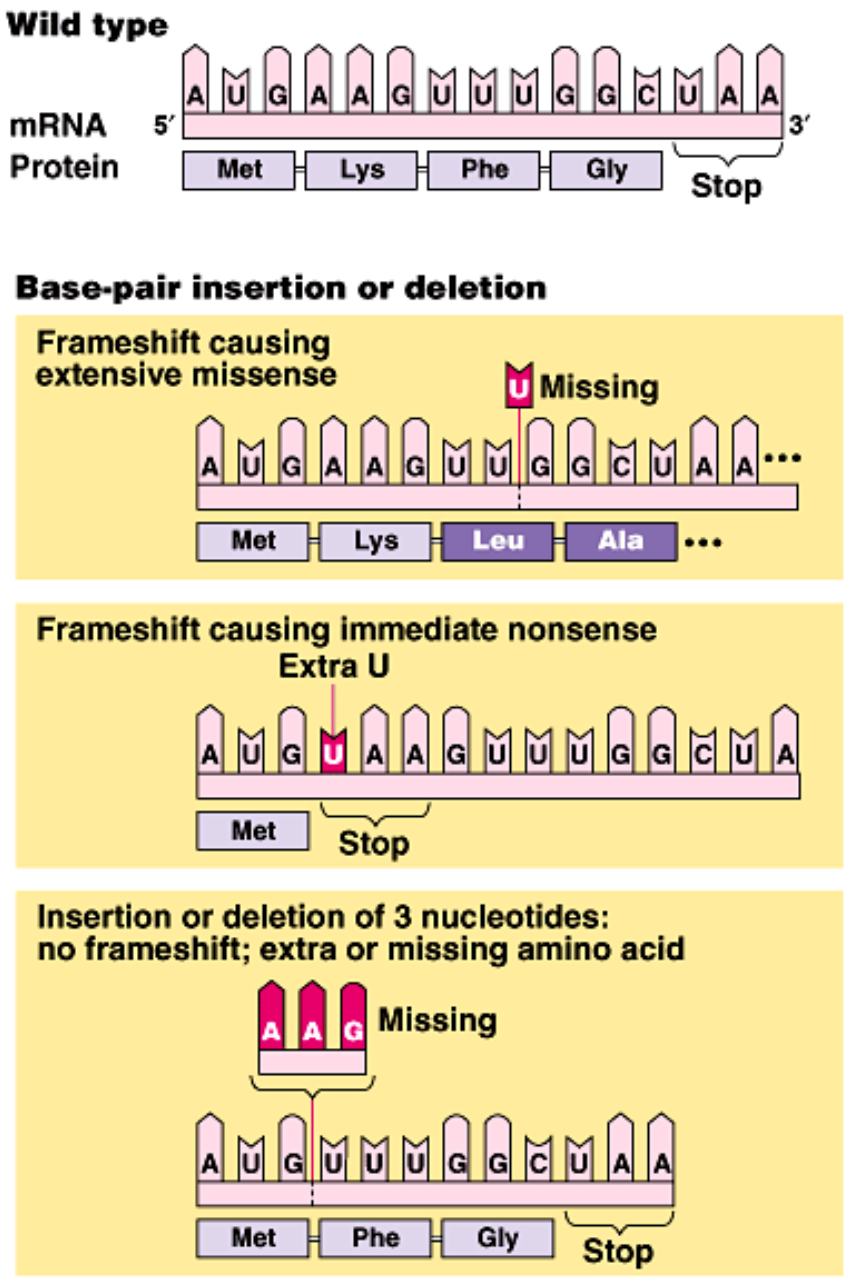
B. Mutations: Heritable Changes in Genes

- Point mutations (silent, missense, nonsense, or frame-shift) result from alterations in single base pairs of DNA.

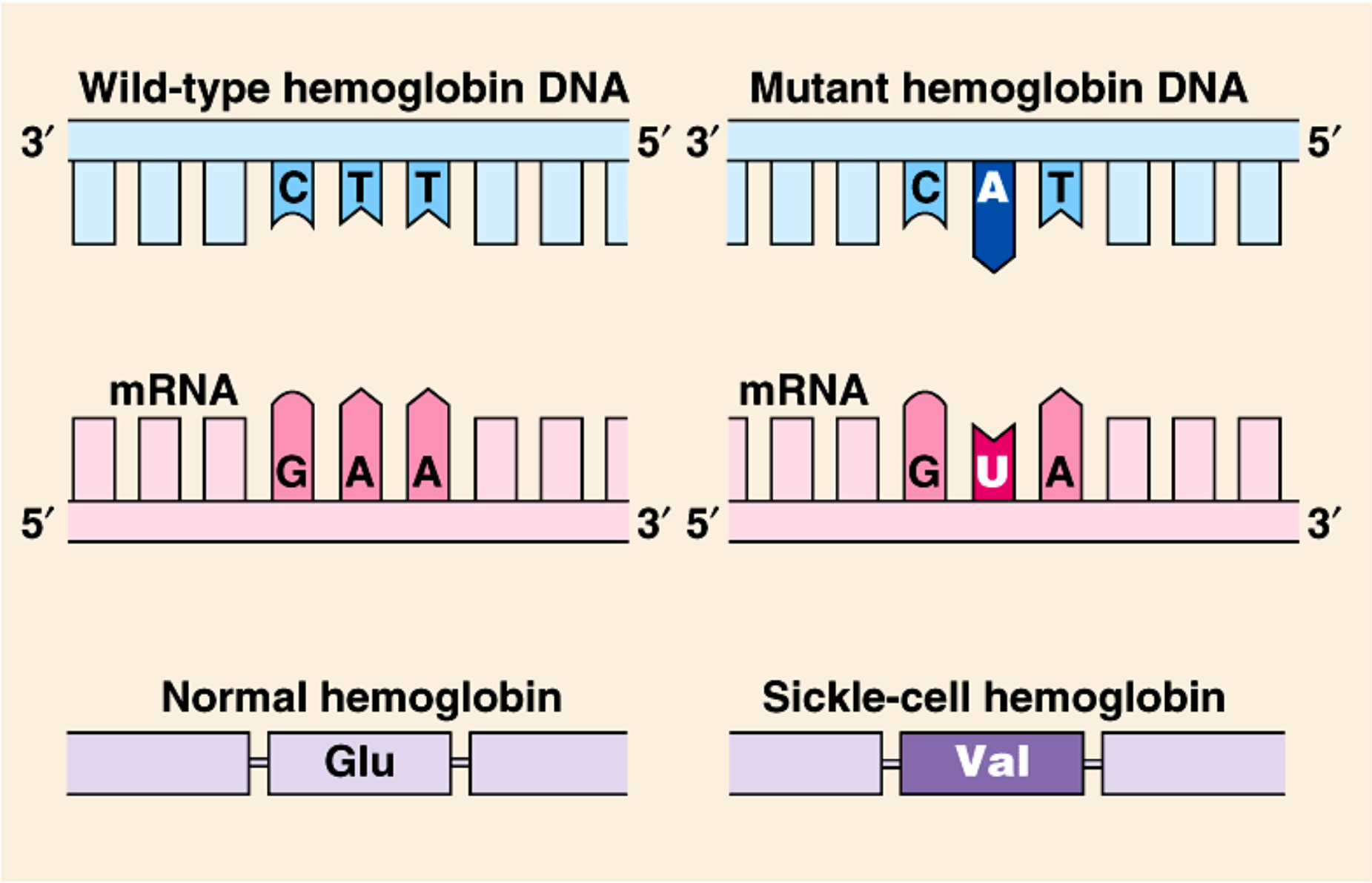
Categories and consequences of point mutations: Base-pair substitution



Categories and consequences of point mutations: Base-pair indels



The molecular basis of sickle-cell disease: a point mutation



B. Mutations: Heritable Changes in Genes

- Chromosomal mutations (deletions, duplications, inversions, or translocations) involve large regions of a chromosome.

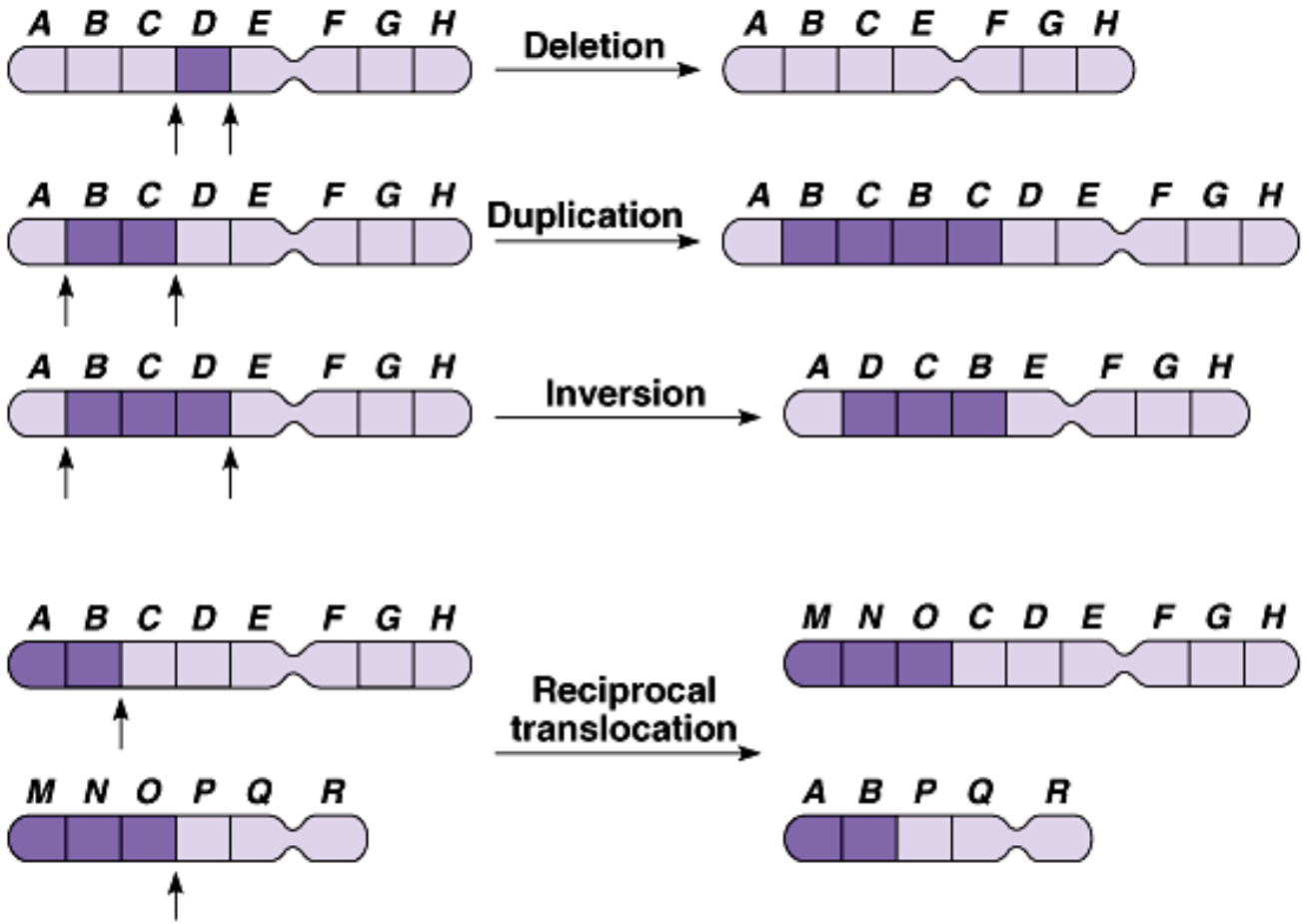
Alterations of chromosome structure

(a) A **deletion** removes a chromosomal segment.

(b) A **duplication** repeats a segment.

(c) An **inversion** reverses a segment within a chromosome.

(d) A **translocation** moves a segment from one chromosome to another, non-homologous one.



C. Repetitive Sequences

- Highly repetitive DNA is present in up to millions of copies of short sequences. It is not transcribed. Its role is unknown.

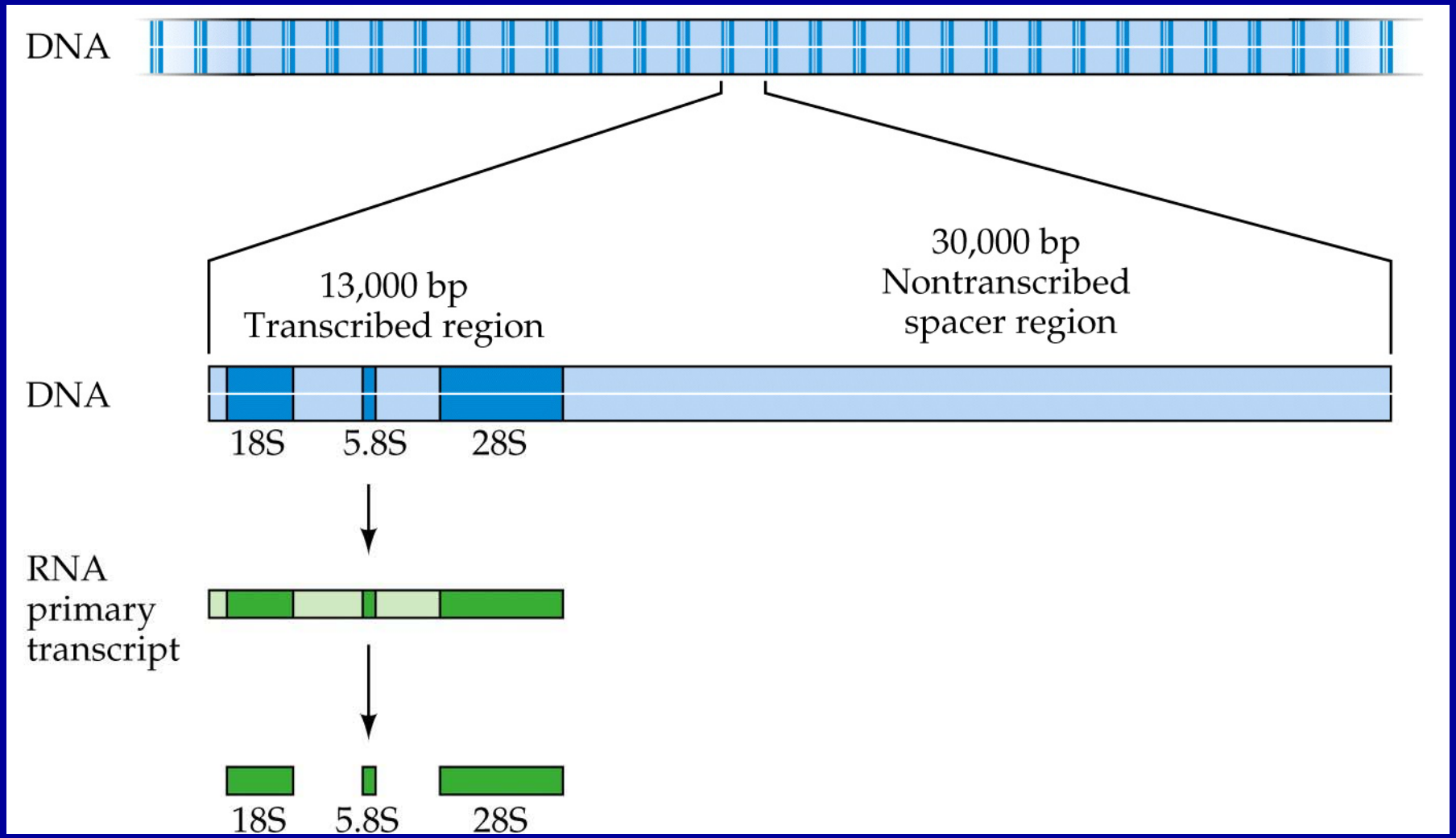
- Rem: Some moderately repetitive DNA sequences, such as telomeric DNA is found at the ends of chromosomes.

C. Repetitive Sequences

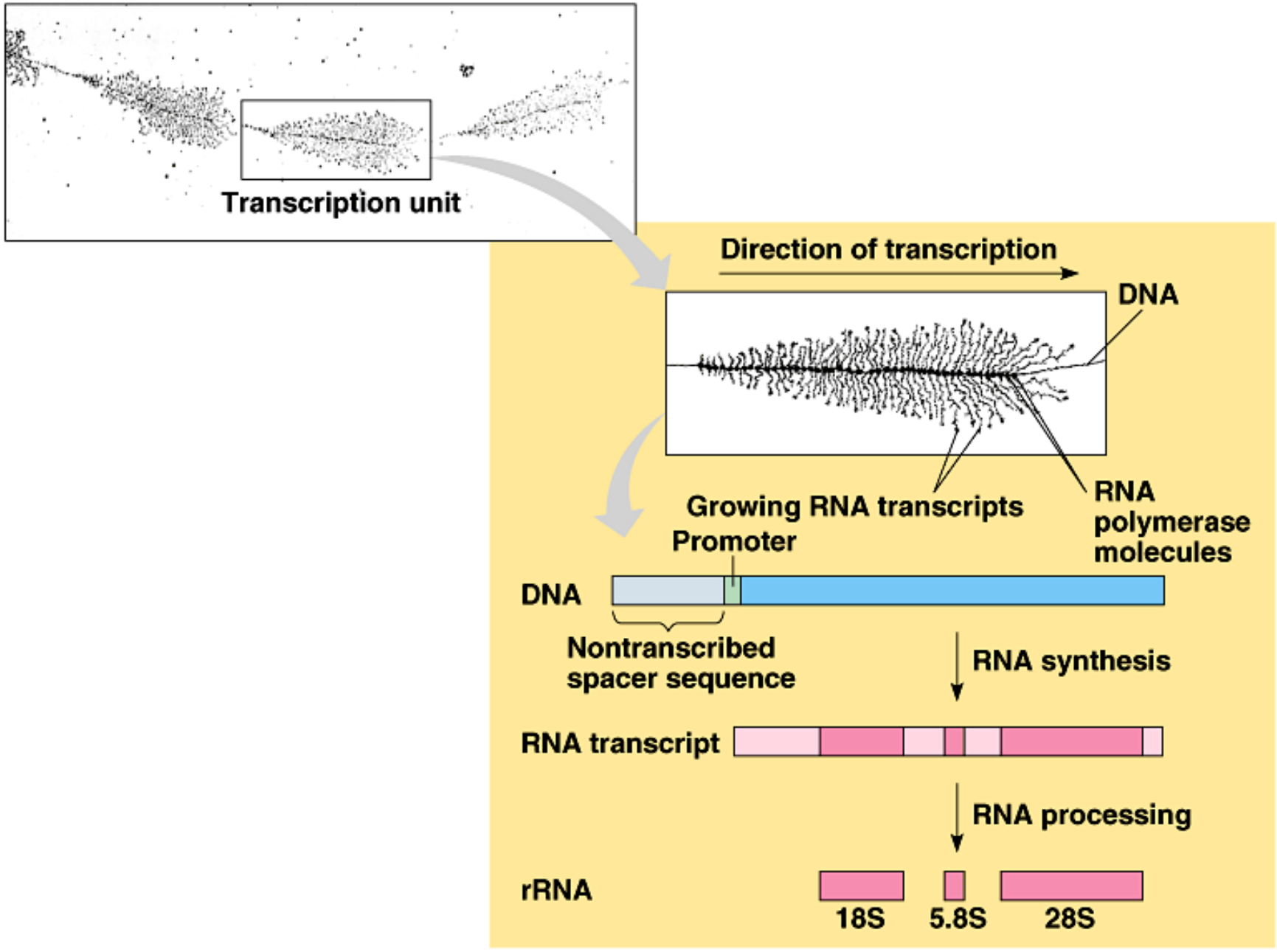
- Some moderately repetitive DNA sequences, such as those coding for ribosomal RNA's, are transcribed.

- Up to three rRNAs result, two go to the large subunit and one goes to the small subunit.

Moderately repetitive DNA sequences

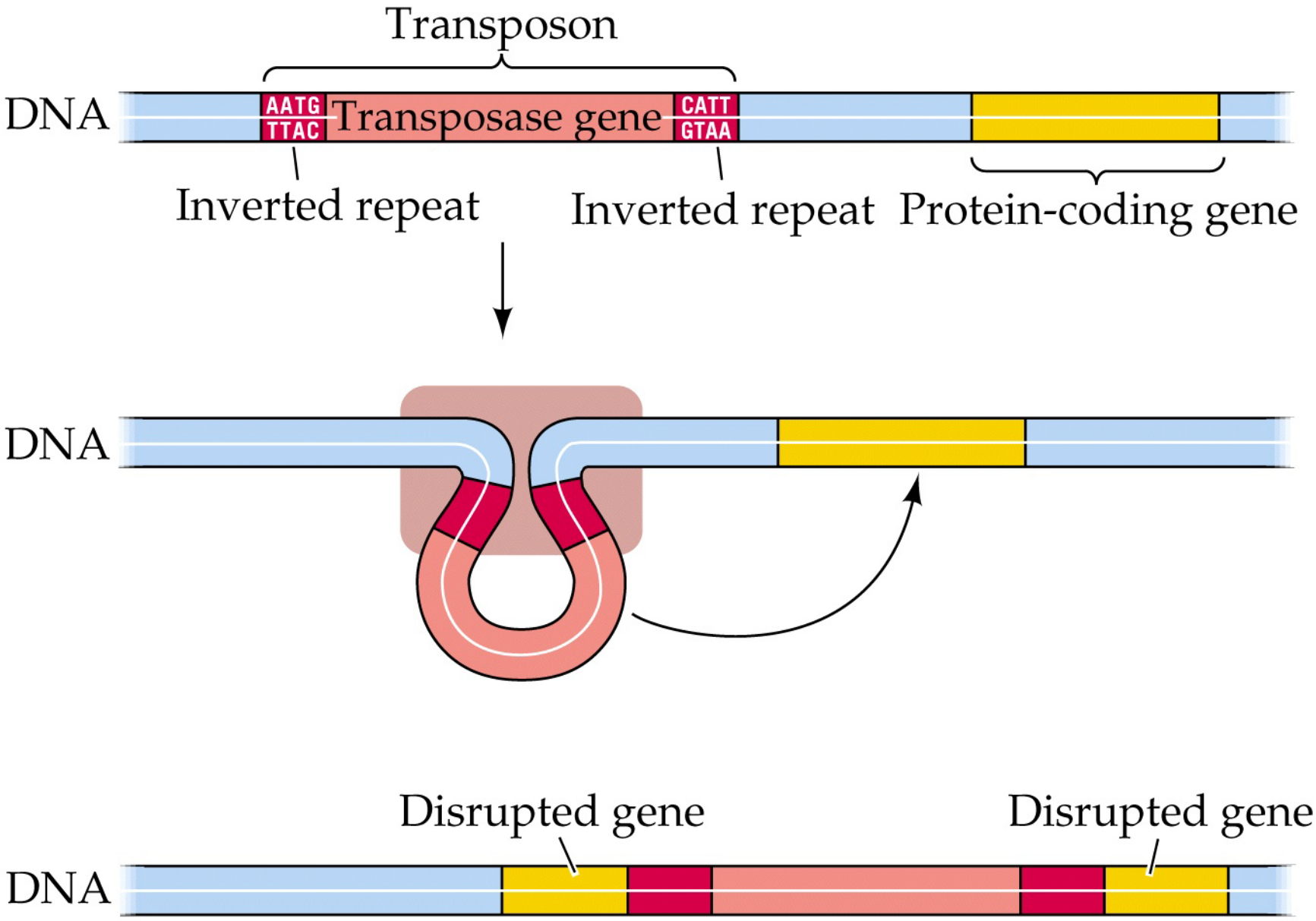


Part of a family of identical genes for ribosomal RNA



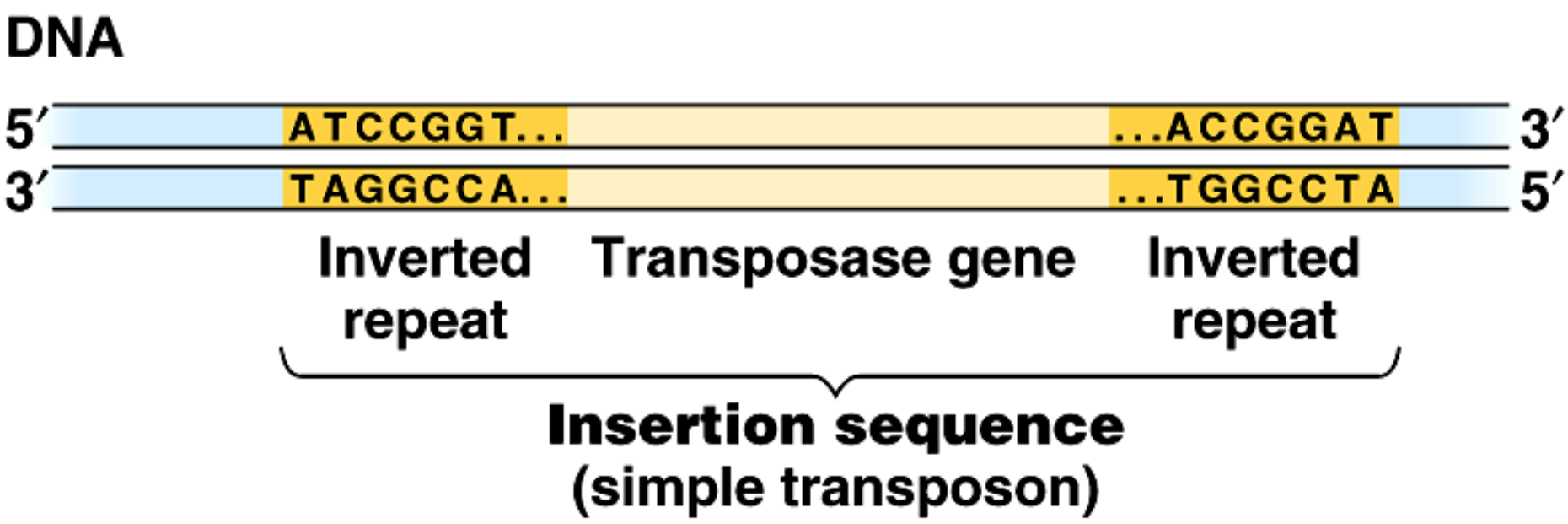
C. Repetitive Sequences

- Some moderately repetitive DNA sequences are transposable, or able to move about the genome. These are known as Transposons.
- Transposons can jump from place to place on the chromosome by actually moving or by making a new copy, inserted at a new location.

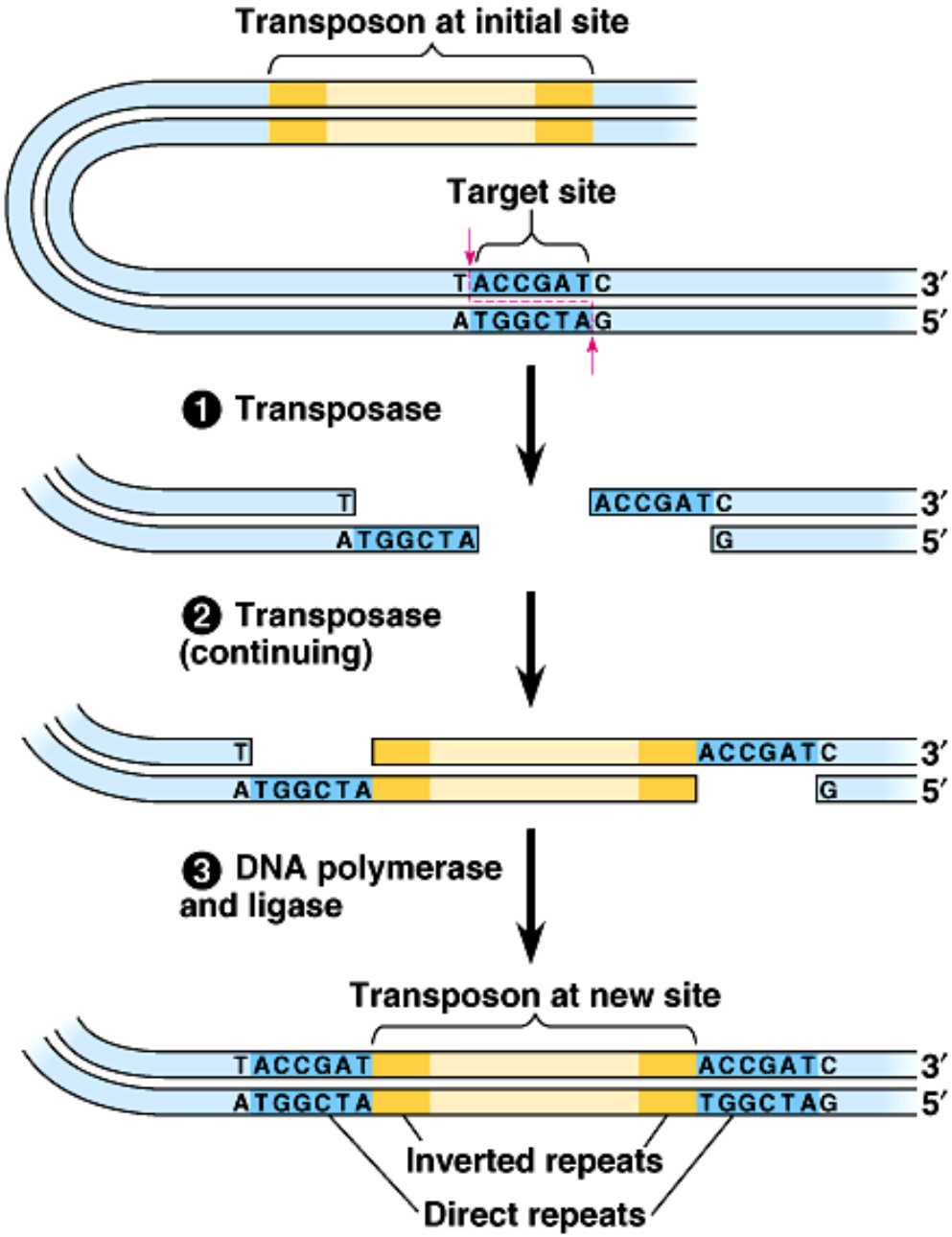

Transposon
DNA
AATG
TTAC
Transposase gene
CATT
GTAA
Inverted repeat
Inverted repeat
Protein-coding gene
DNA
Disrupted gene
Disrupted gene
DNA

Transposons in corn

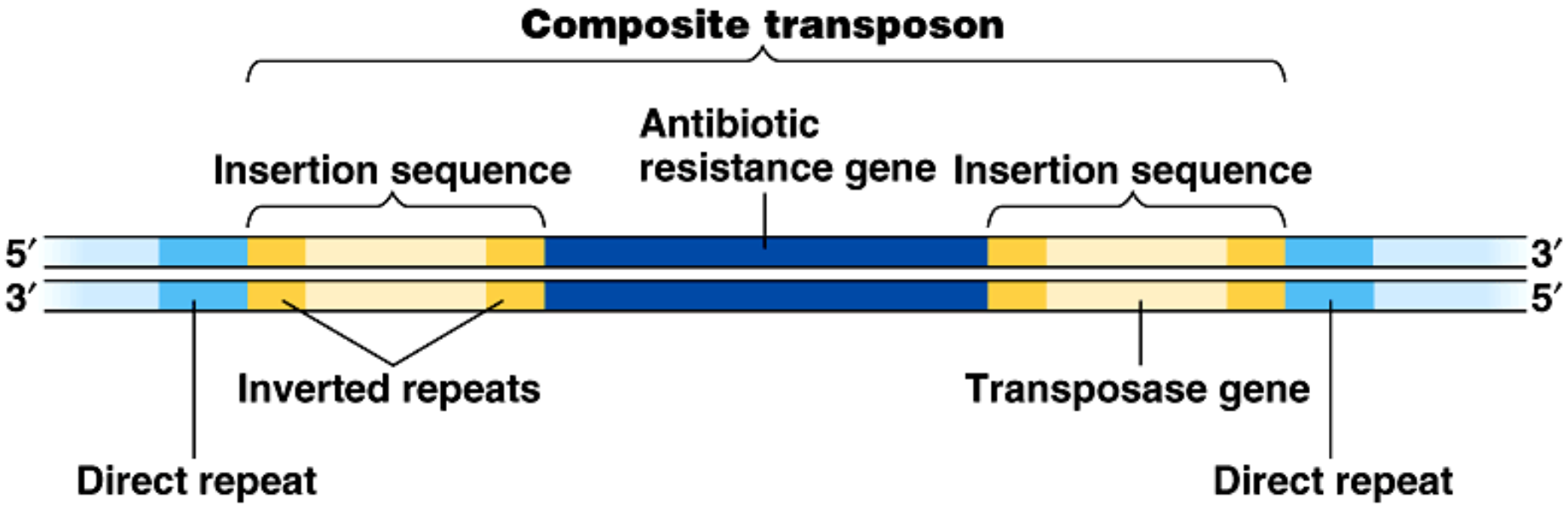
Insertion sequences, the simplest transposons



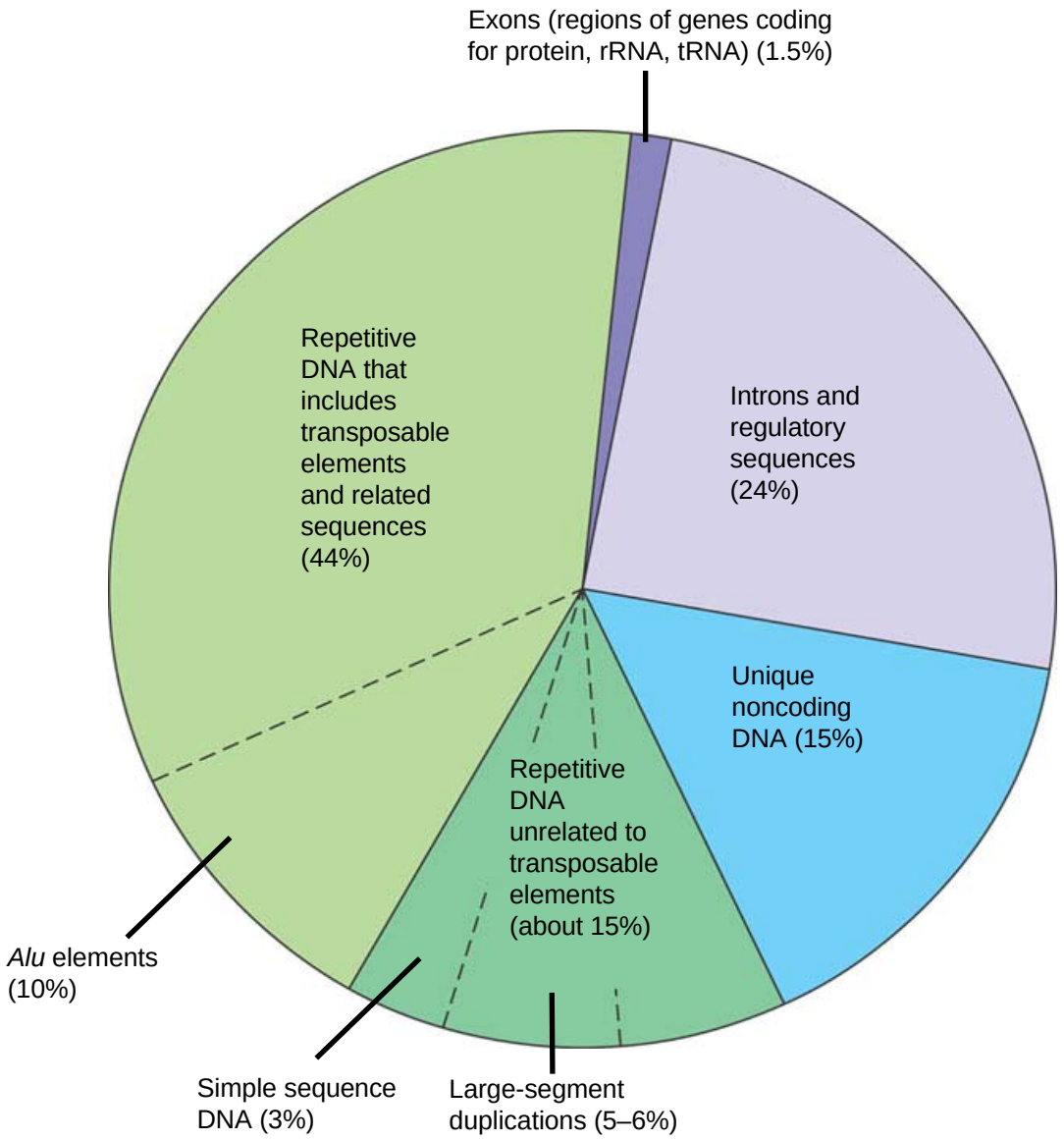
Insertion of a transposon and creation of direct repeats



Anatomy of a composite transposon

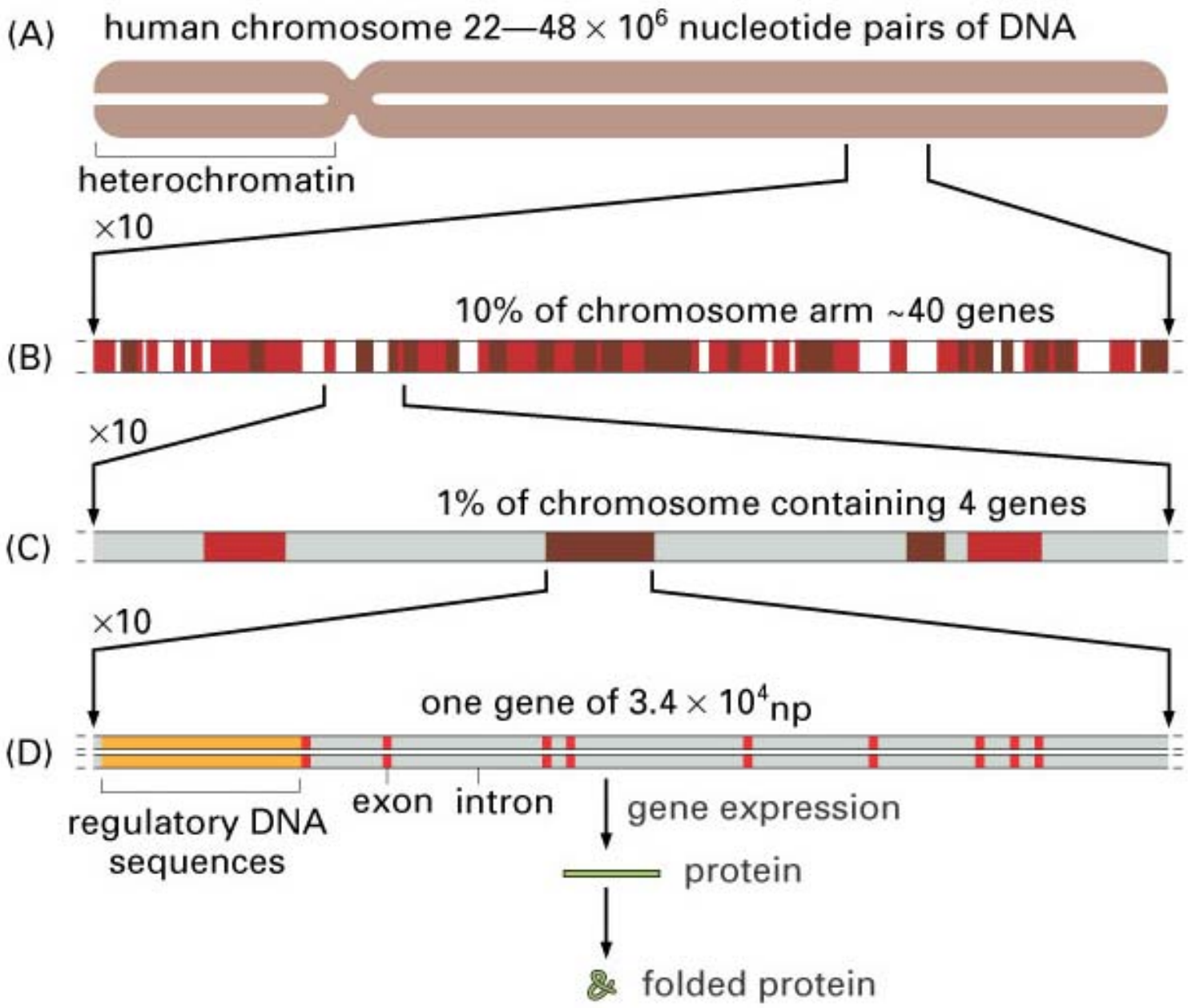


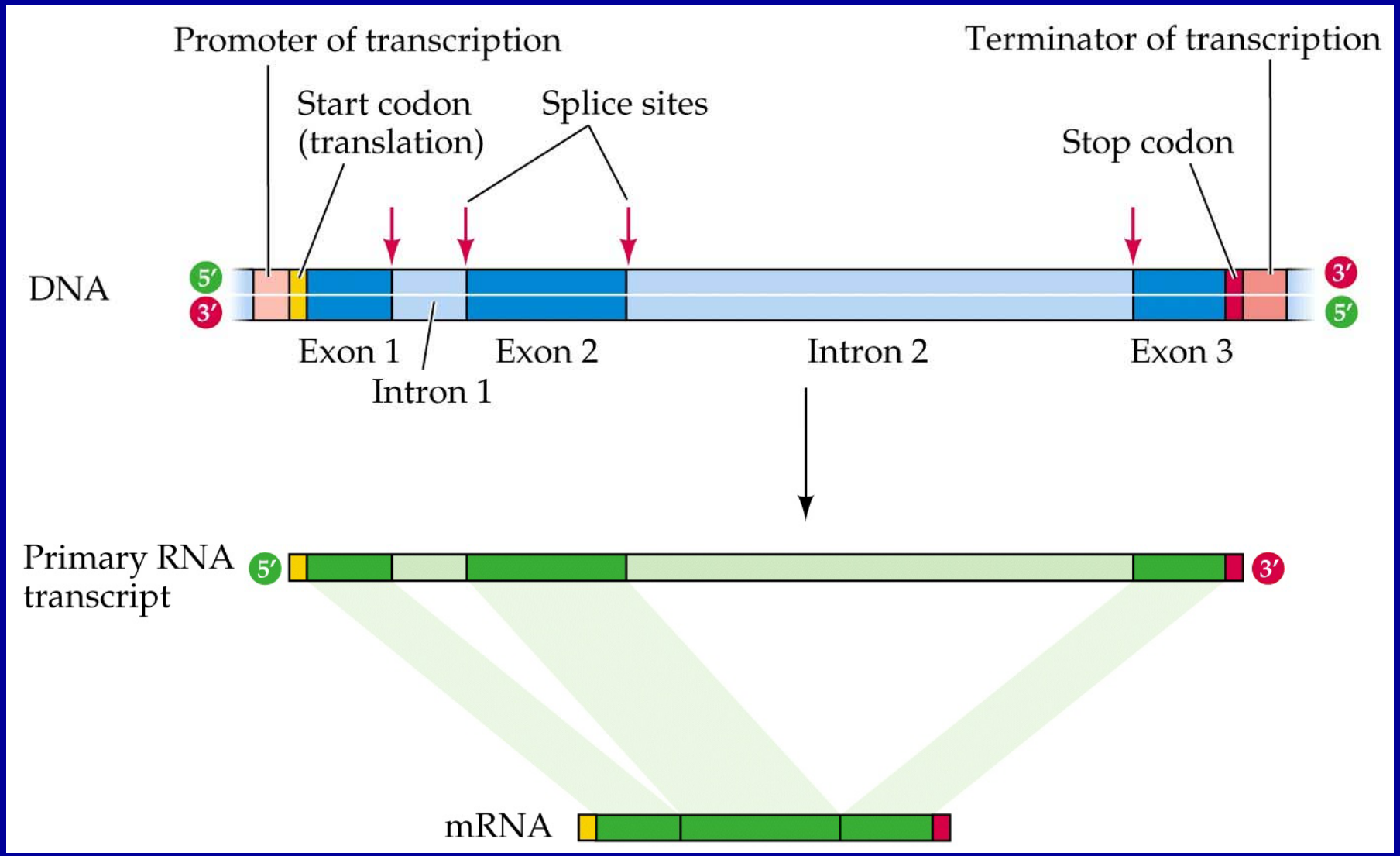
Types of DNA sequences in the human genome



D. The Structures of Protein-Coding Genes

- A typical protein-coding gene has noncoding internal sequences (introns) as well as flanking sequences that are involved in the machinery of transcription and translation in addition to its exons or coding regions.
- These are usually single copy genes.

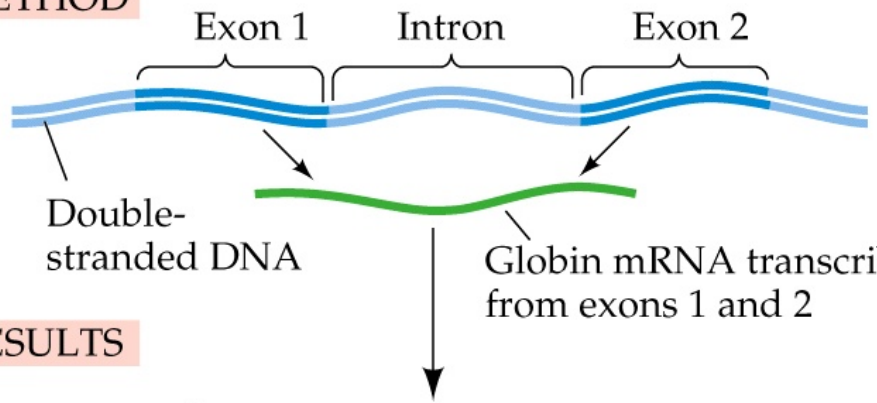

(A)
human chromosome 22—48 × 10^6 nucleotide pairs of DNA
heterochromatin
×10
10% of chromosome arm ~40 genes
(B)
×10
1% of chromosome containing 4 genes
(C)
×10
one gene of 3.4 × 10^4 np
(D)
regulatory DNA sequences
exon
intron
gene expression
protein
& folded protein


Promoter of transcription
Terminator of transcription
Start codon (translation)
Splice sites
Stop codon
DNA
5′
3′
3′
5′
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
Primary RNA transcript
5′
3′
mRNA

EXPERIMENT

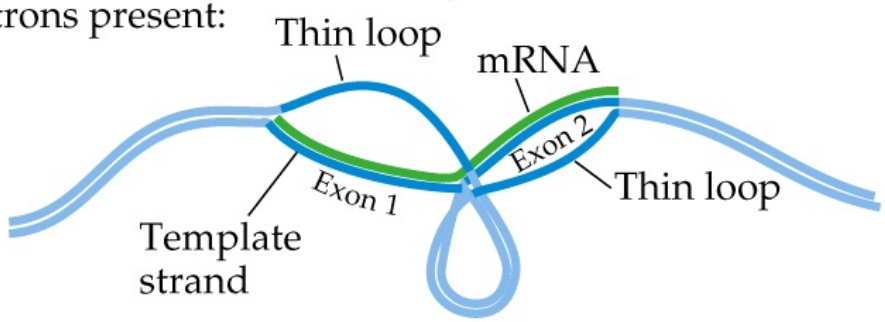
Question: Are there regions within the coding sequence of a gene that do not end up in its mRNA?

METHOD

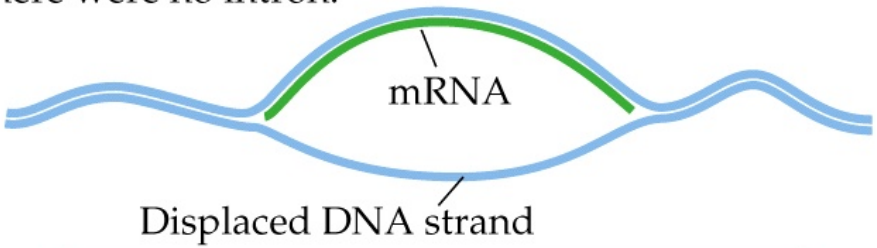


RESULTS

Introns present:



If there were no intron:

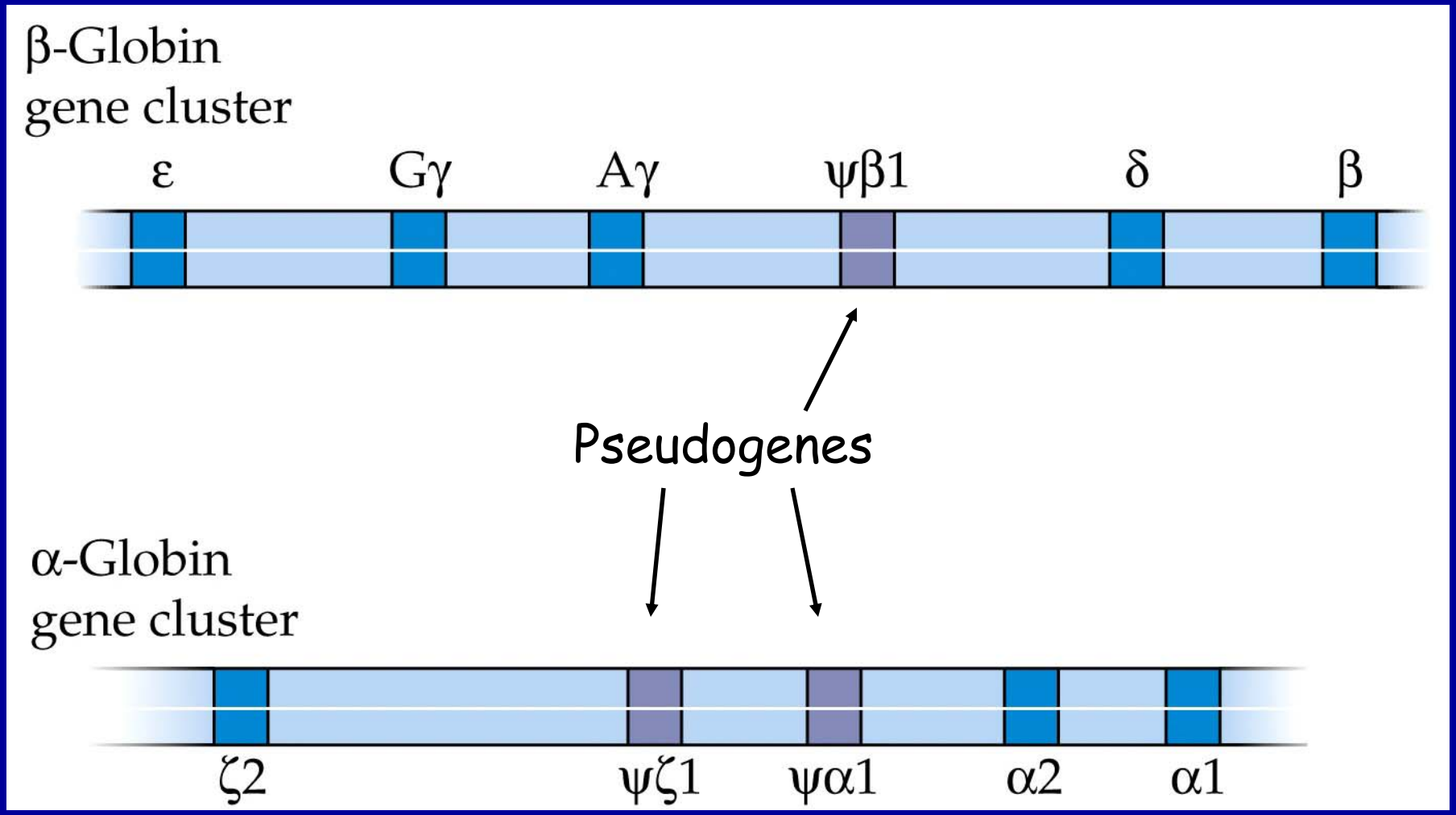


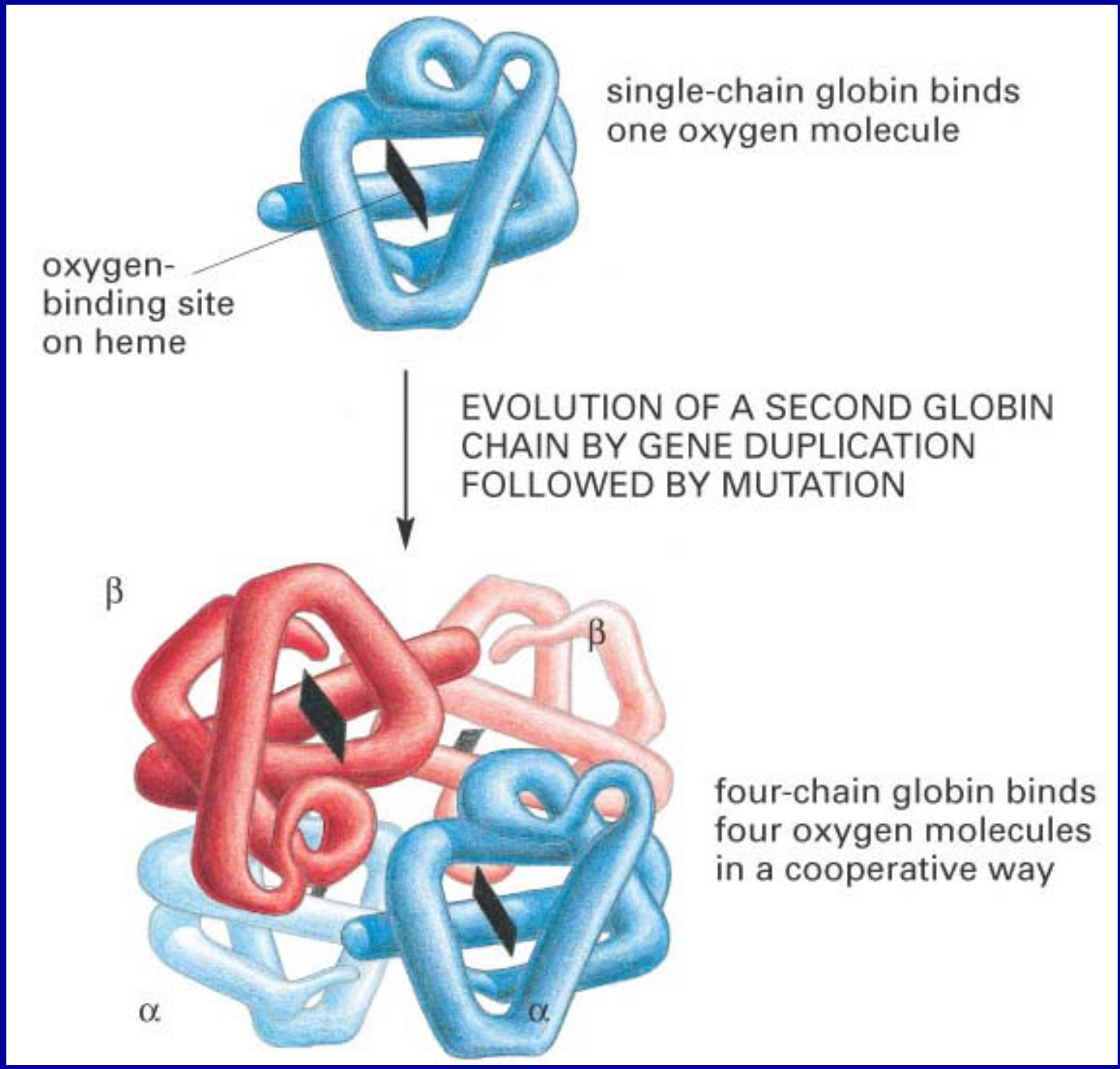
Conclusion: The final mRNA does not contain noncoding internal regions in a gene in DNA.

D. The Structures of Protein-Coding Genes

- Some eukaryotic genes form families of related genes that have similar sequences and code for similar proteins. These related proteins may be made at different times and in different tissues.
- Some sequences in gene families are pseudogenes, which code for nonfunctional mRNA's or proteins.

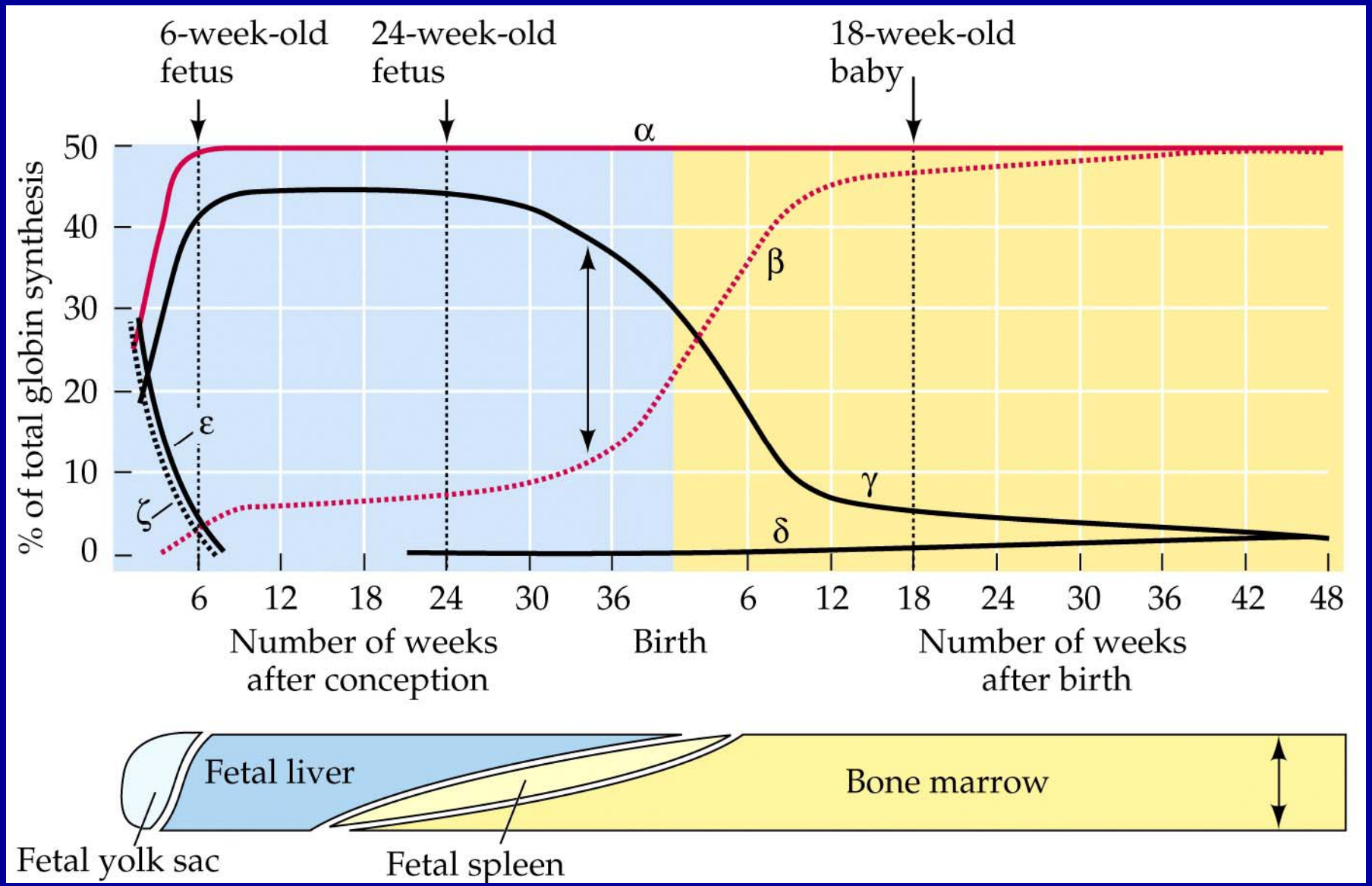
Gene Families



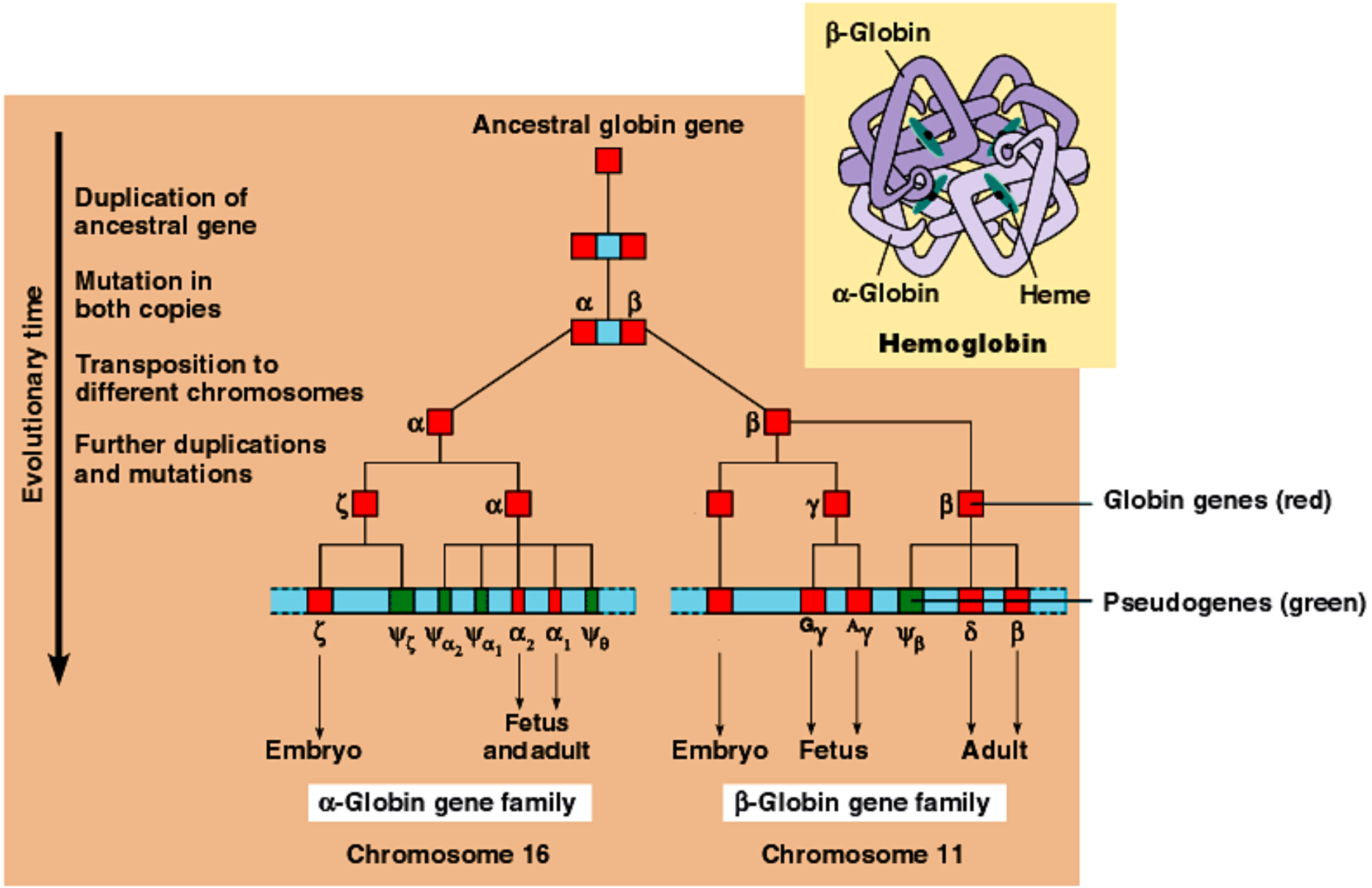

single-chain globin binds
one oxygen molecule
oxygen-
binding site
on heme
EVOLUTION OF A SECOND GLOBIN
CHAIN BY GENE DUPLICATION
FOLLOWED BY MUTATION
β
β
α
α
four-chain globin binds
four oxygen molecules
in a cooperative way

D. The Structures of Protein-Coding Genes

- Differential expression of different genes in the β-globin family ensures important physiological changes during human development.

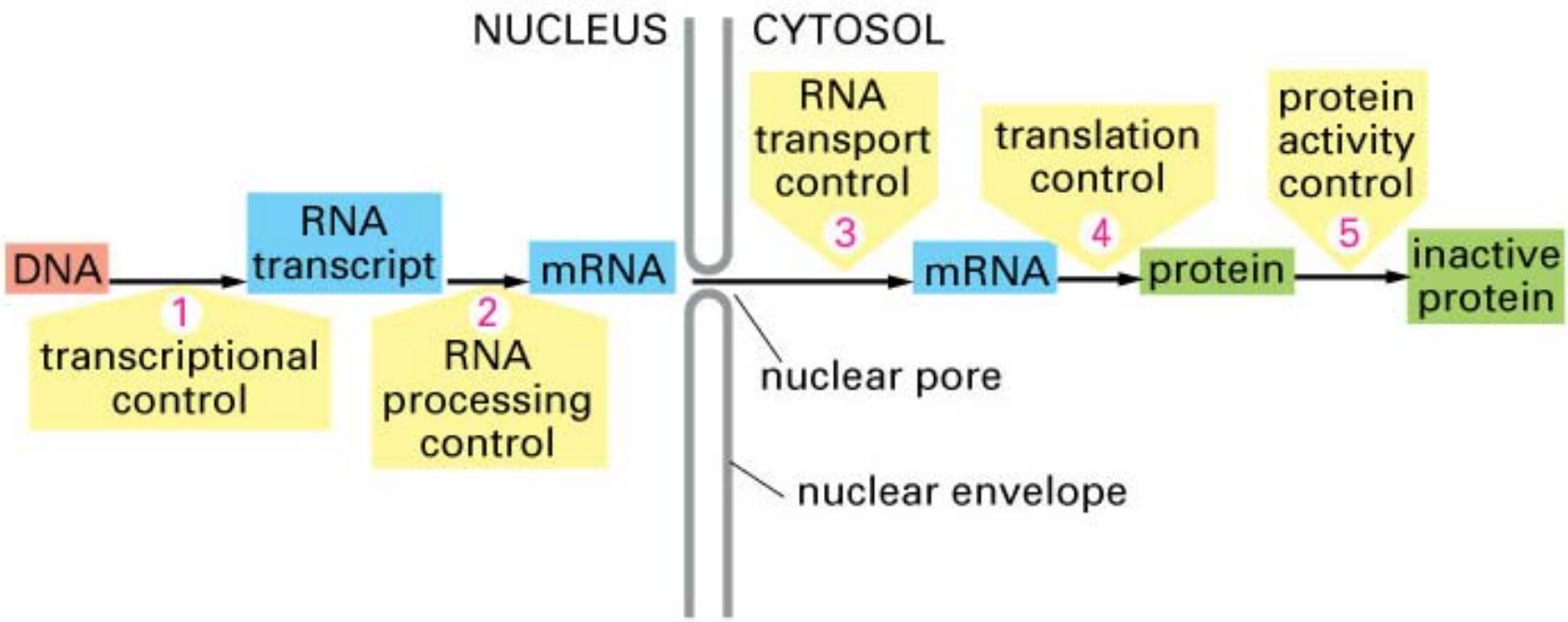

6-week-old fetus
24-week-old fetus
18-week-old baby
α
β
γ
δ
ε
ζ
% of total globin synthesis
50
40
30
20
10
0
6
12
18
24
30
36
6
12
18
24
30
36
42
48
Number of weeks after conception
Birth
Number of weeks after birth
Fetal liver
Bone marrow
Fetal yolk sac
Fetal spleen

The evolution of human α-globin and β-globin gene families



E. Transcriptional Control

- Eukaryotic gene expression can be controlled at the transcriptional, posttranscriptional, translational, and posttranslational levels.


NUCLEUS
CYTOSOL
RNA transport control
3
translation control
4
protein activity control
5
DNA
RNA transcript
mRNA
mRNA
protein
inactive protein
1
transcriptional control
2
RNA processing control
nuclear pore
nuclear envelope

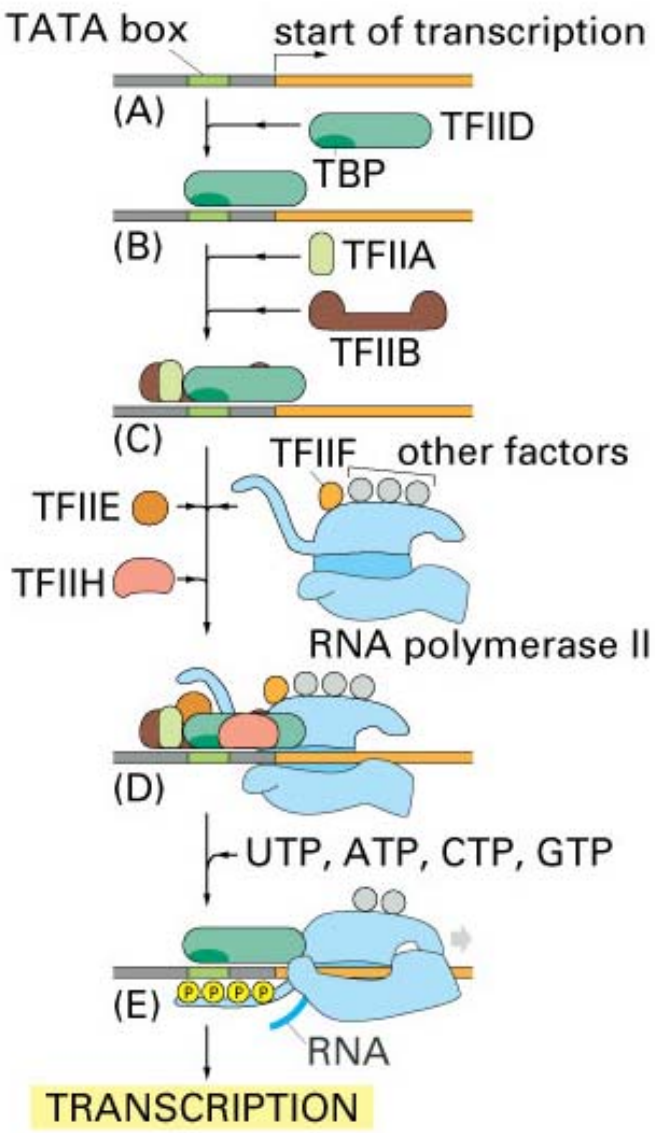
E. Transcriptional Control

- The major method of control of eukaryotic gene expression is selective transcription, which results from specific proteins binding to regulatory regions on DNA.

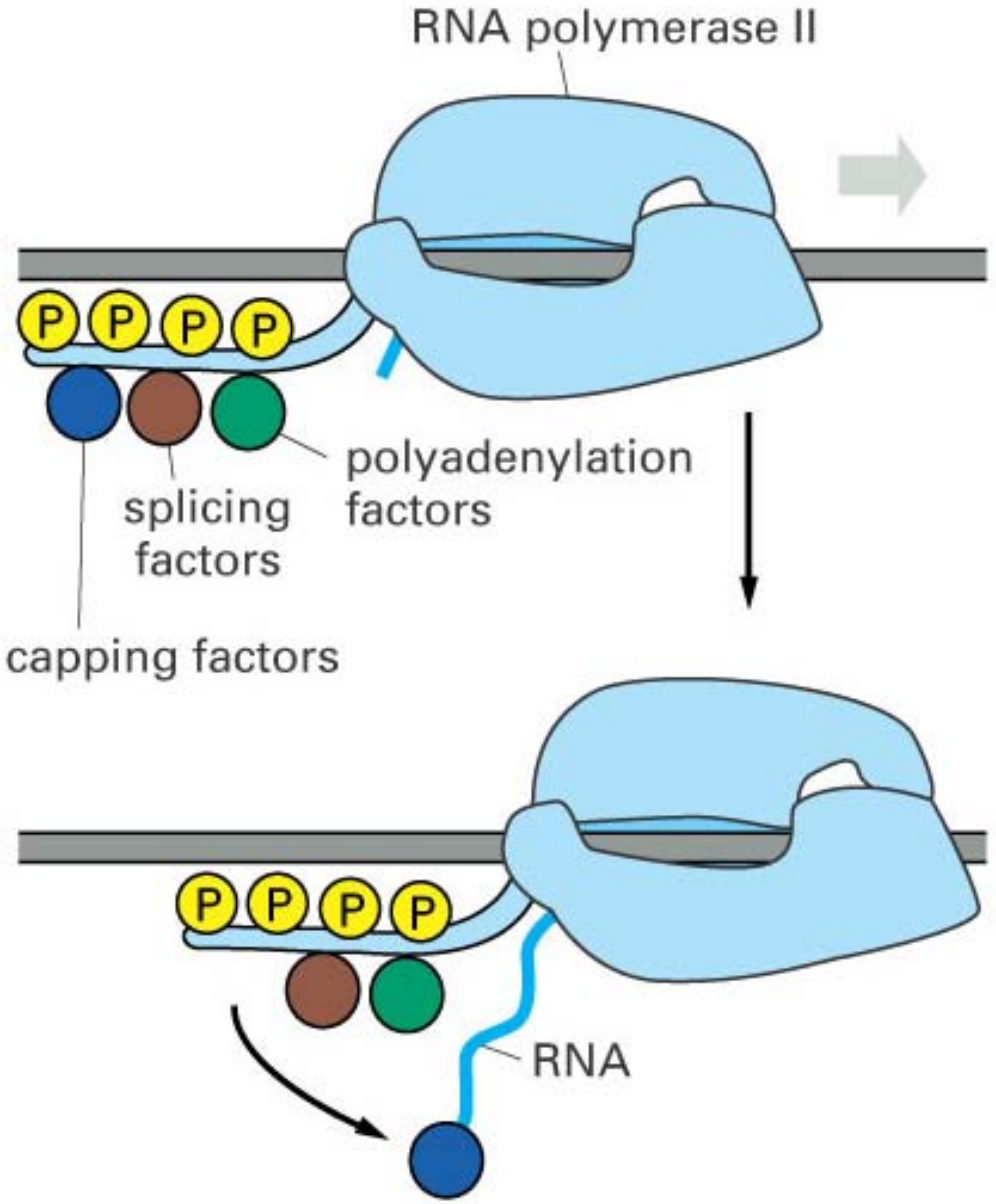
E. Transcriptional Control

- A series of "general" transcription factors must bind to the promoter before RNA polymerase can bind.
- Whether RNA polymerase will initiate transcription also depends on the binding of regulatory proteins, activator proteins, and repressor proteins.

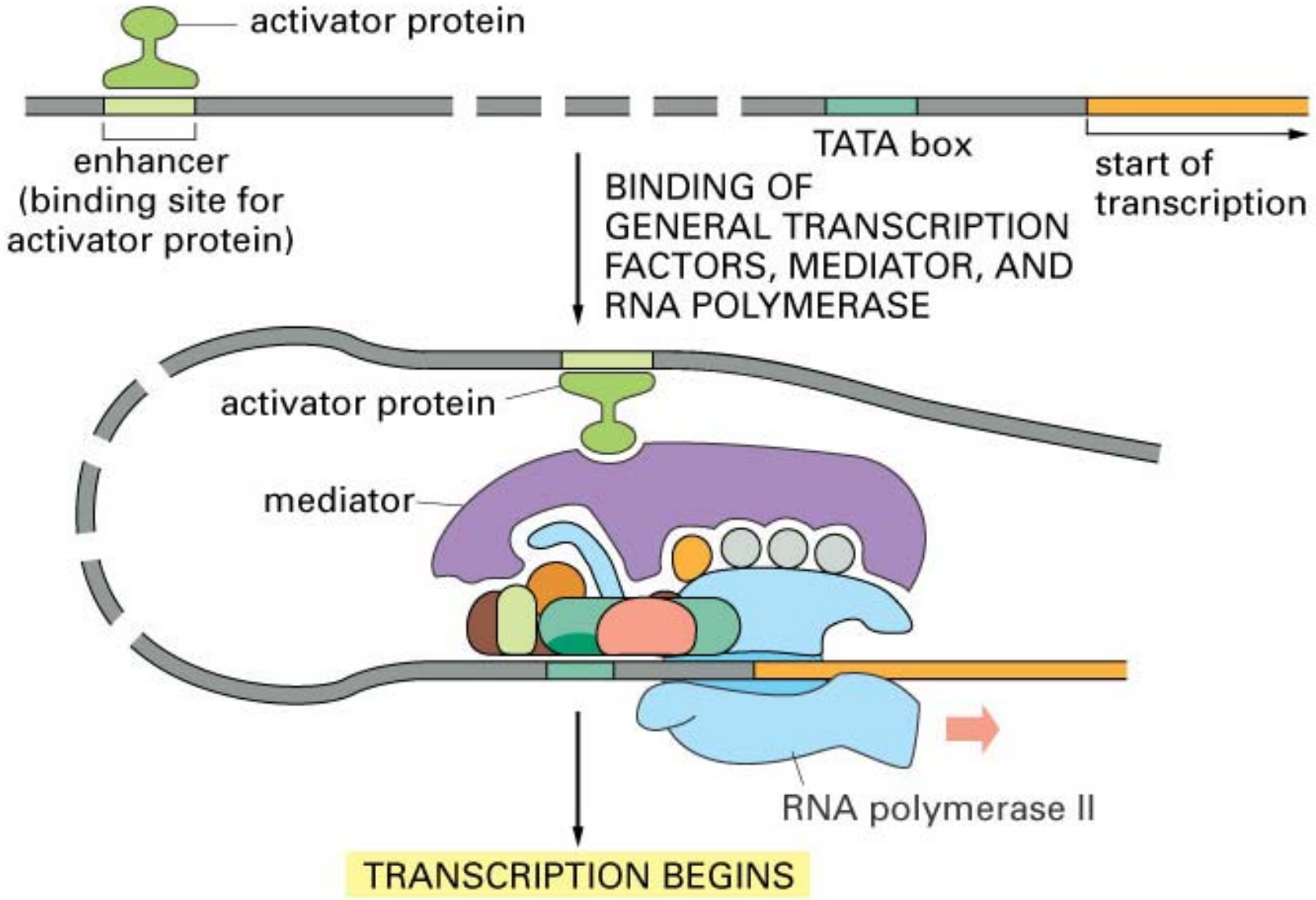
RNA pol II requires many "general" transcription factors



Phosphorylation of RNA pol II allows RNA processing proteins to ride on its tail



Action of distal enhancers and transcription activators

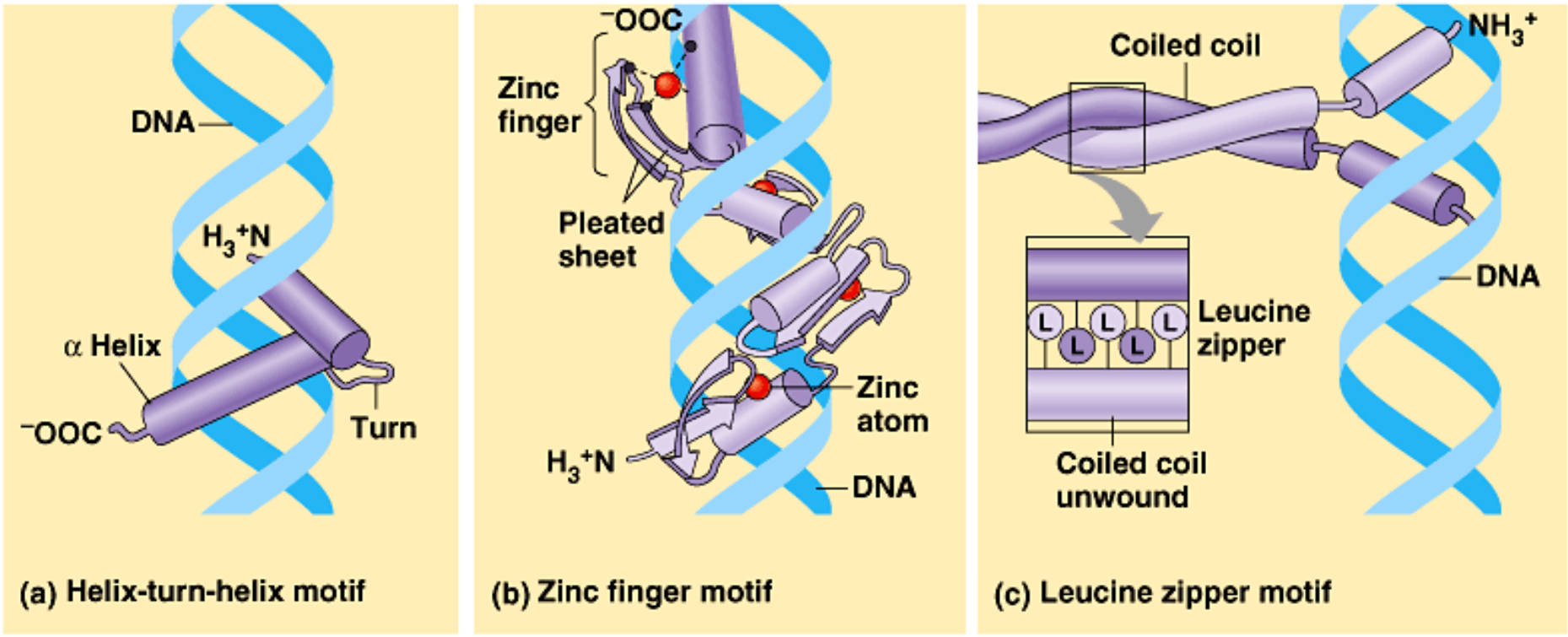


Repressors/Silencers too!

E. Transcriptional Control

- The DNA-binding domains of most DNA-binding proteins have one of four structural motifs: helix-turn-helix, zinc finger, leucine zipper, or homeodomain.

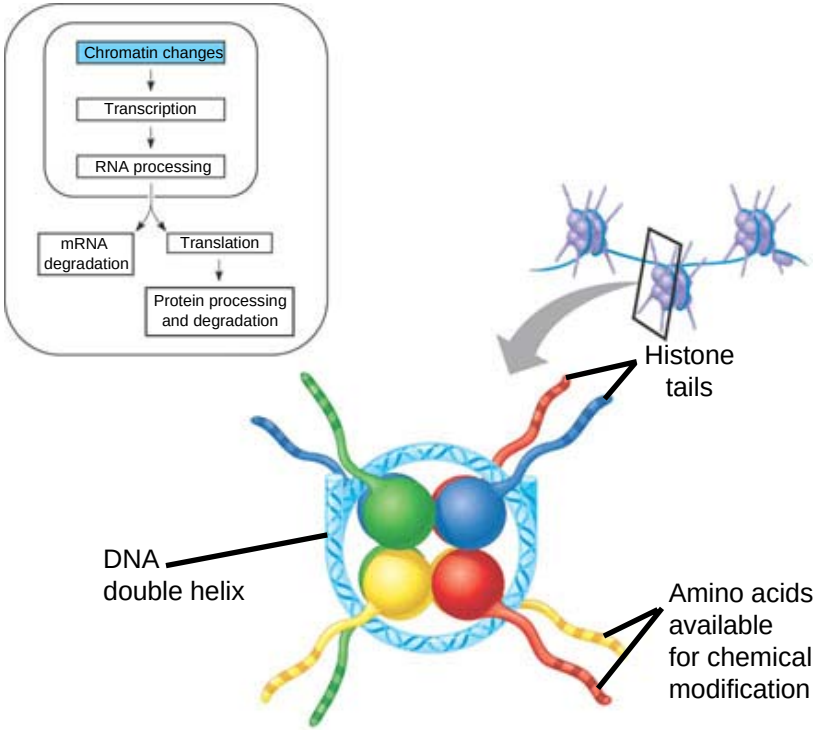
Three of the major types of DNA-binding domains in transcription factors



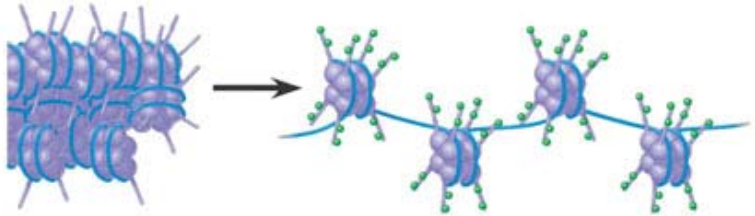
E. Transcriptional Control

- Acetylation of histone tails promotes loose chromatin structure that permits transcription to more readily occur.

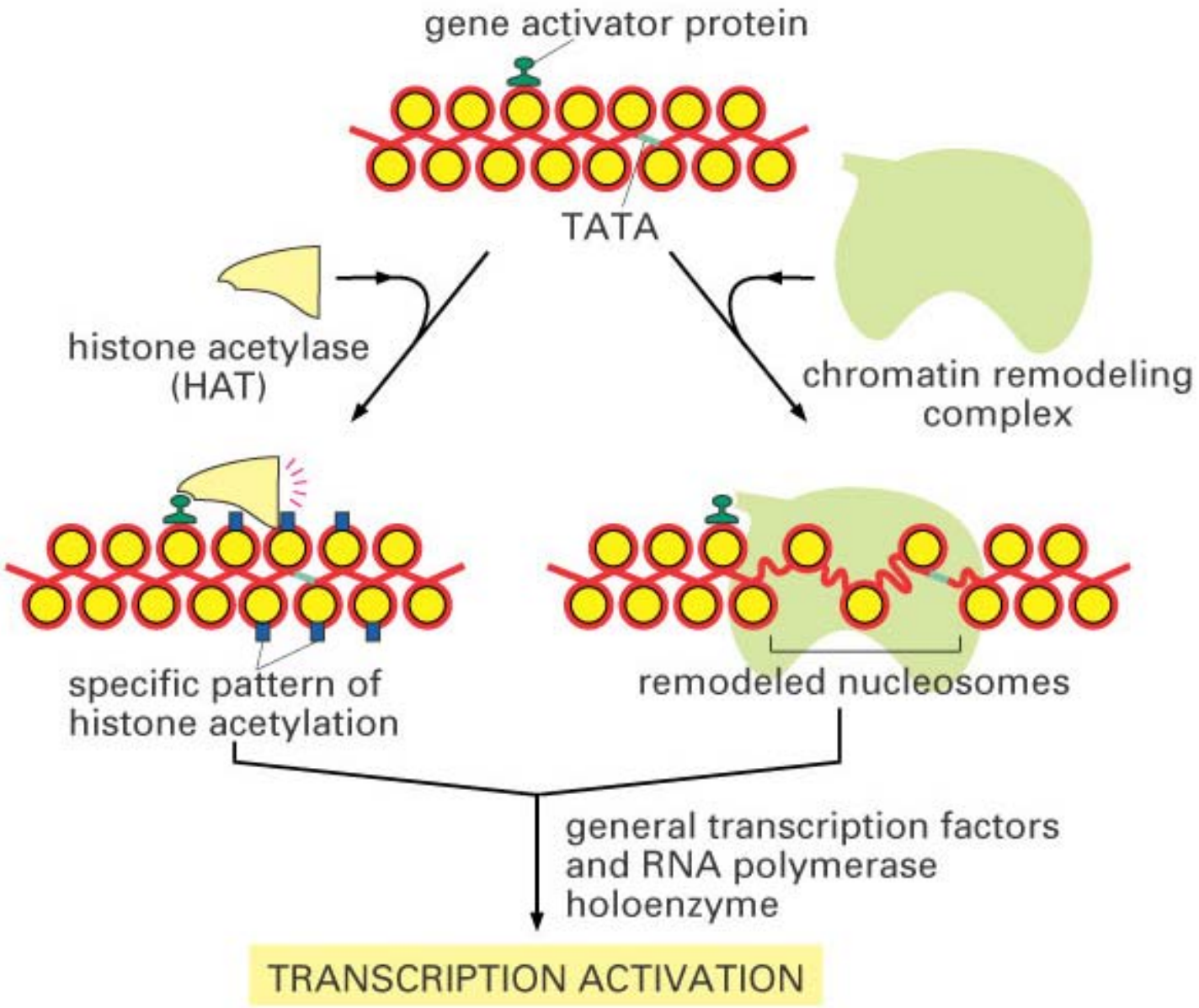
A simple model of histone tails and the effect of histone acetylation



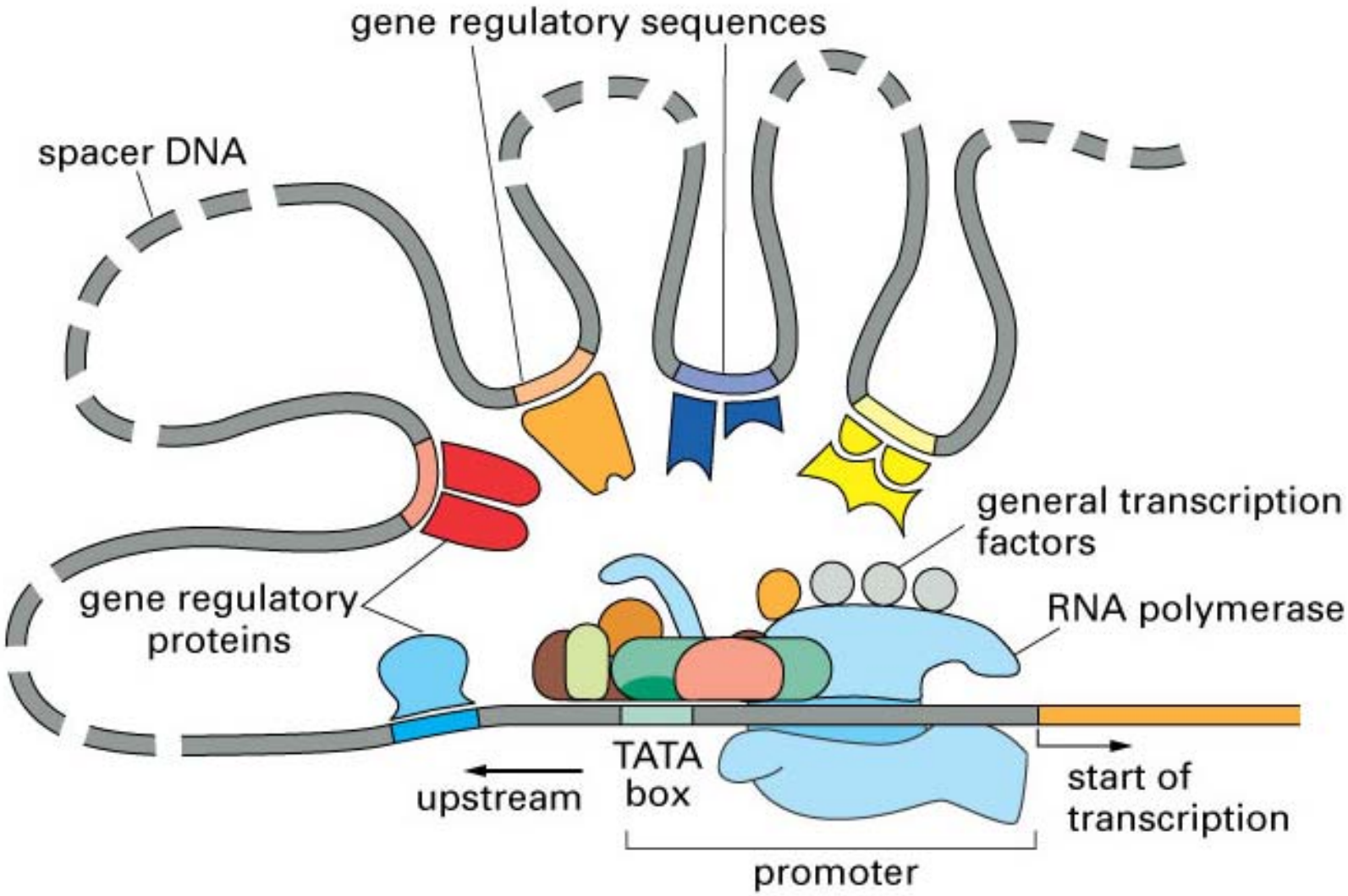
(a) Histone tails protrude outward from a nucleosome



(b) Acetylation of histone tails promotes loose chromatin structure that permits transcription


gene activator protein
TATA
histone acetylase
(HAT)
chromatin remodeling
complex
specific pattern of
histone acetylation
remodeled nucleosomes
general transcription factors
and RNA polymerase
holoenzyme
TRANSCRIPTION ACTIVATION

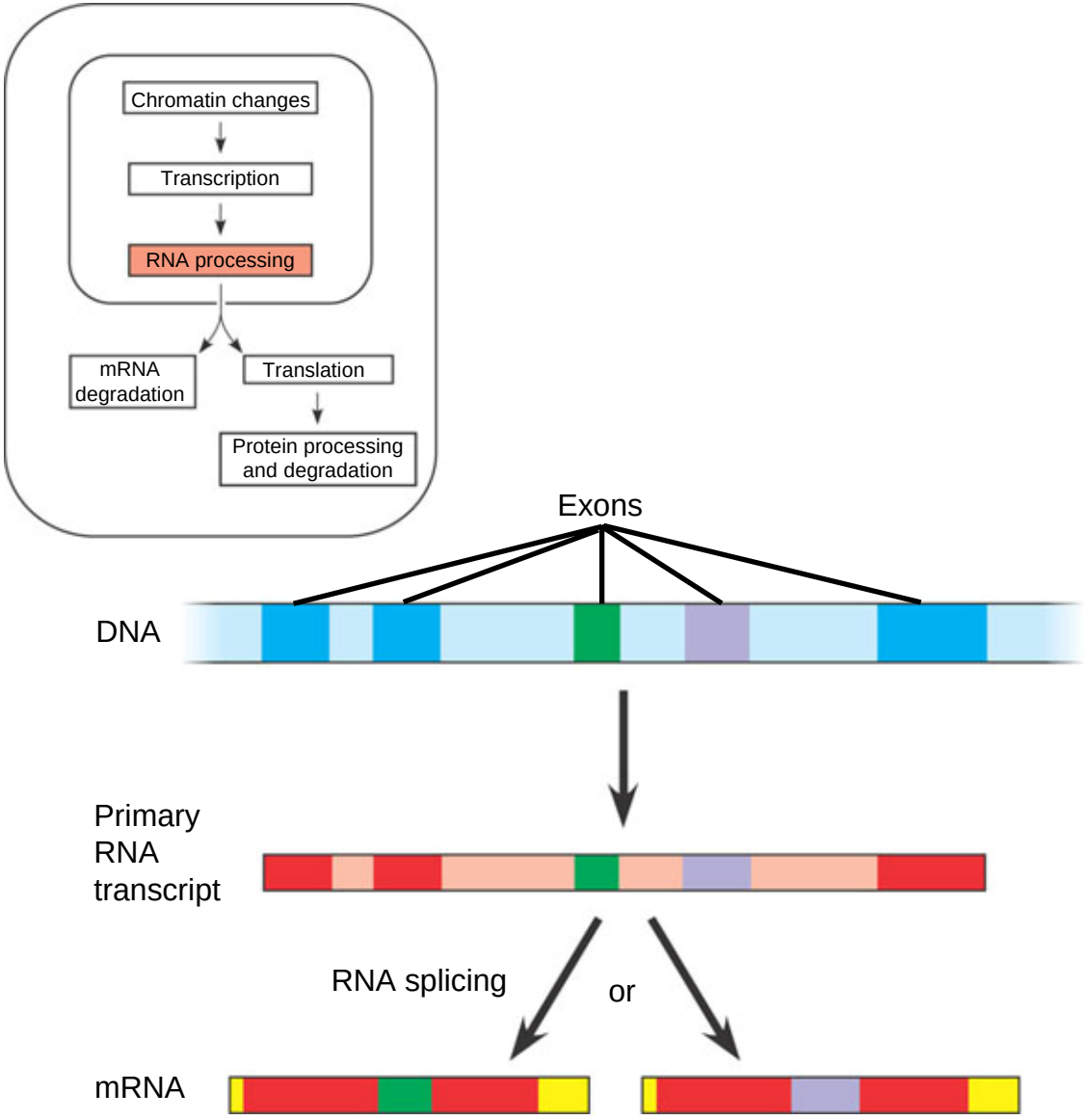
Combinatorial control regulation concept



F. Posttranscriptional Control

- Because eukaryotic genes have several exons, alterative mRNAs can be generated from the same RNA transcript.
- This alternate splicing can be used to produce different proteins.
- The stability of mRNA in the cytoplasm can be regulated by the binding of proteins.

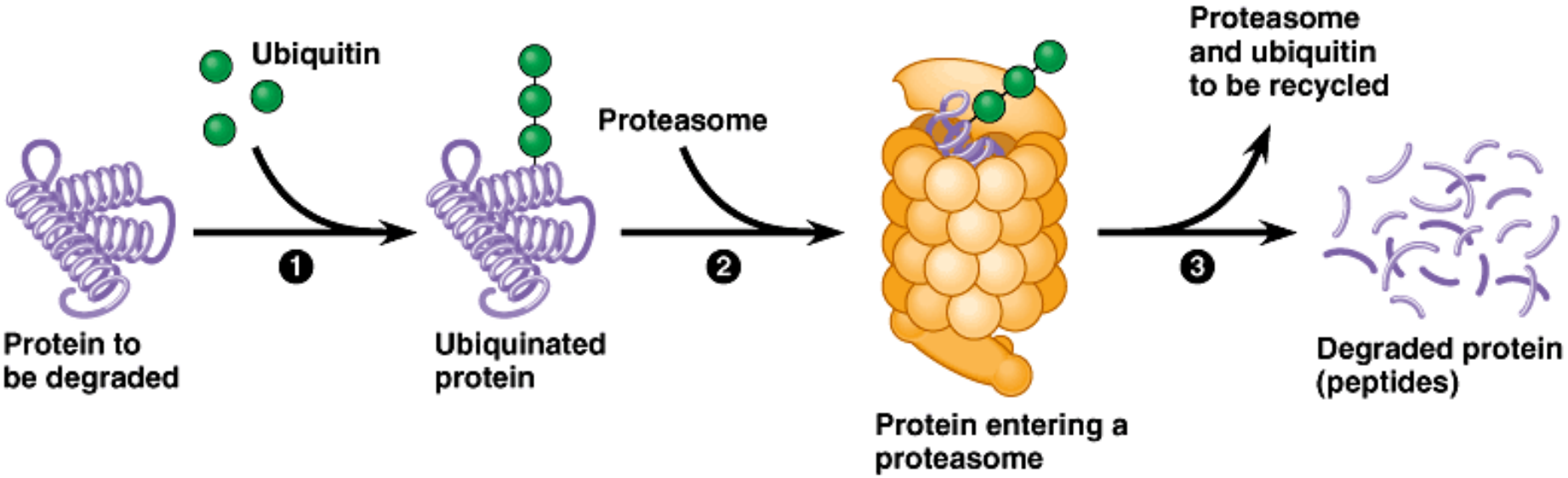
Alternative RNA splicing



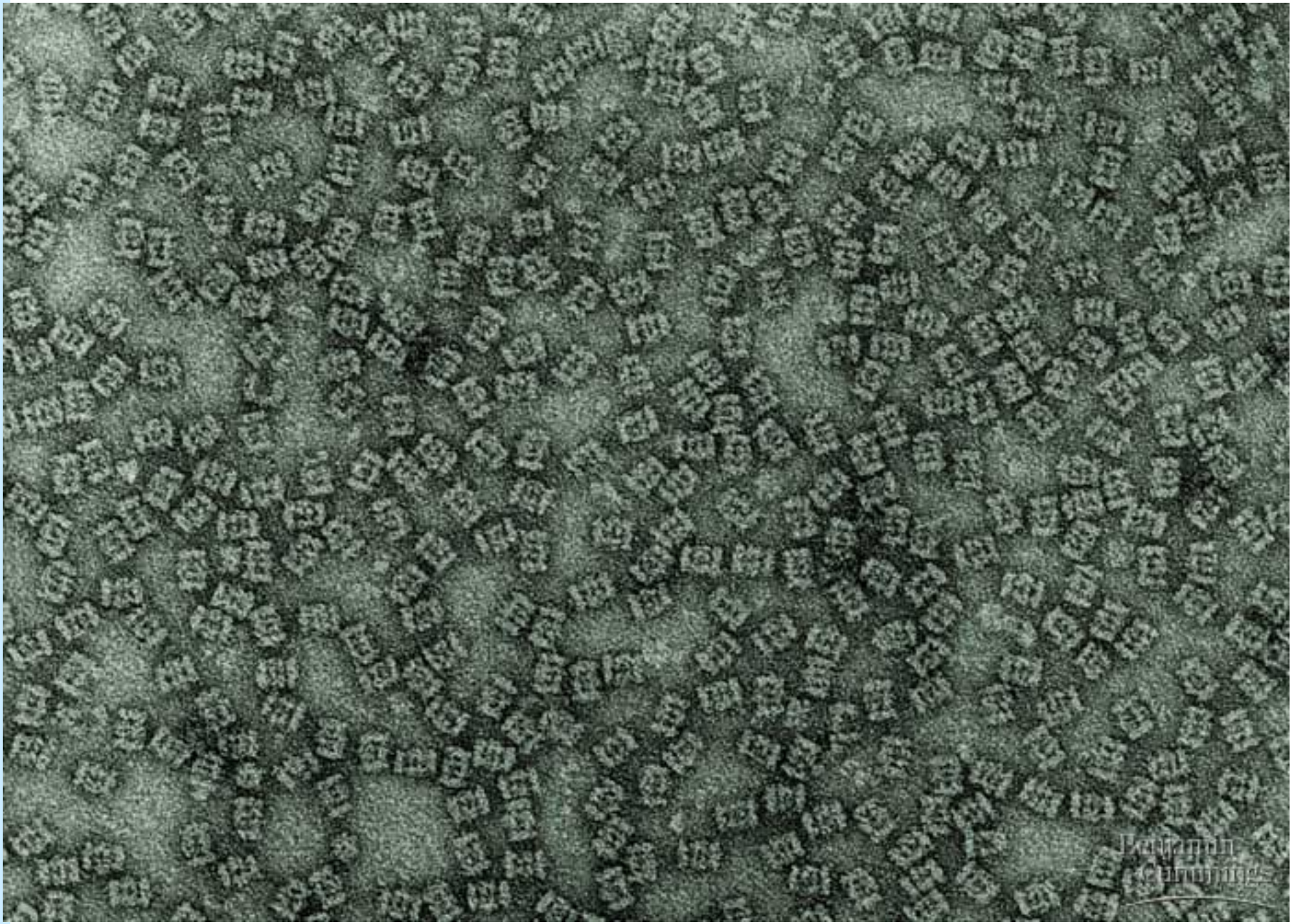
F. Posttranslational Control

- Proteasomes degrade proteins targeted for breakdown.

Degradation of a protein by a proteasome



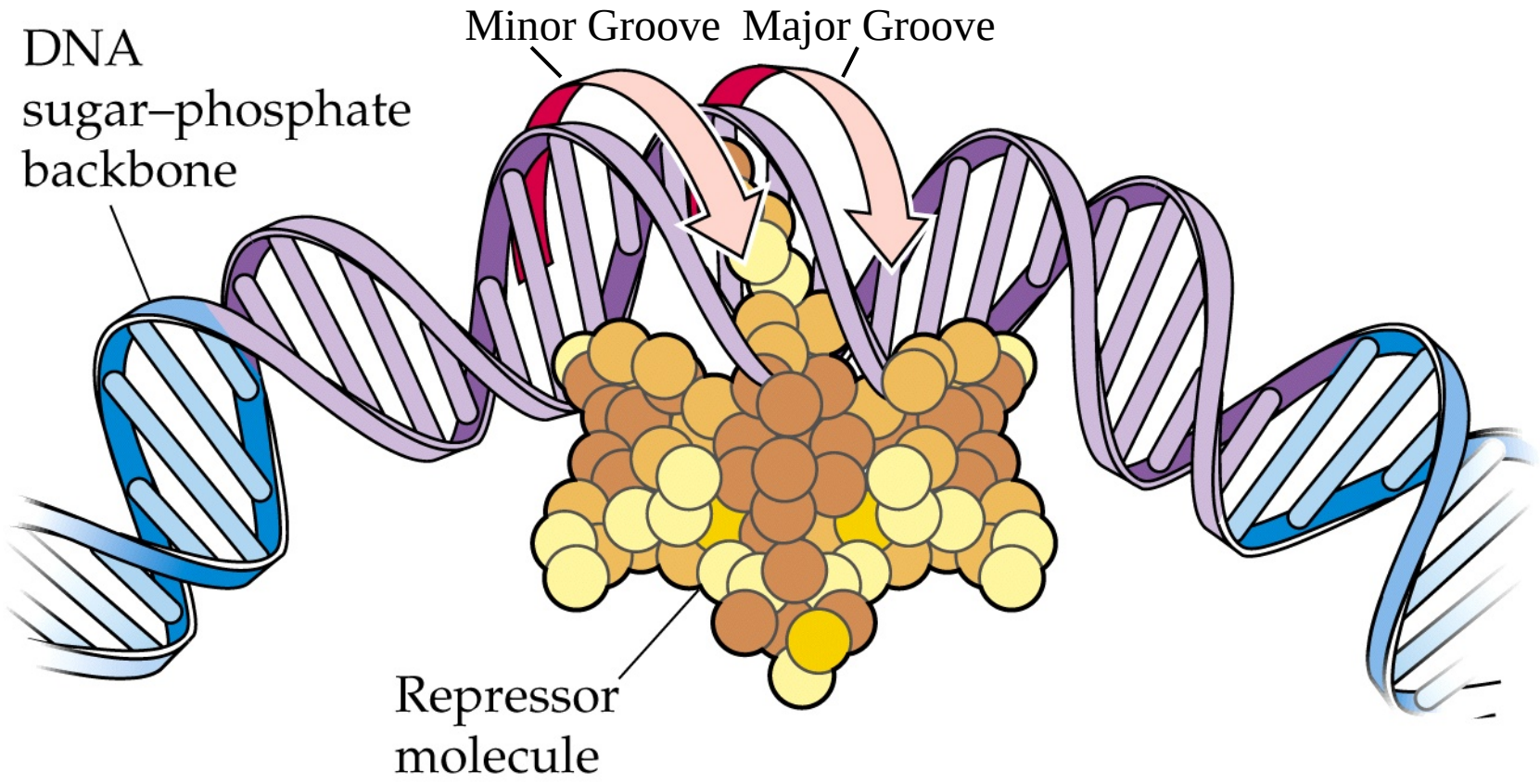
Proteasomes



G. Regulation of Gene Expression in Prokaryotes

- An operon consists of a promoter, an operator, and structural genes. Promoters and operators do not code for proteins, but serve as binding sites for regulatory proteins.
- When a repressor protein binds to the operator, transcription of the structural genes is inhibited.

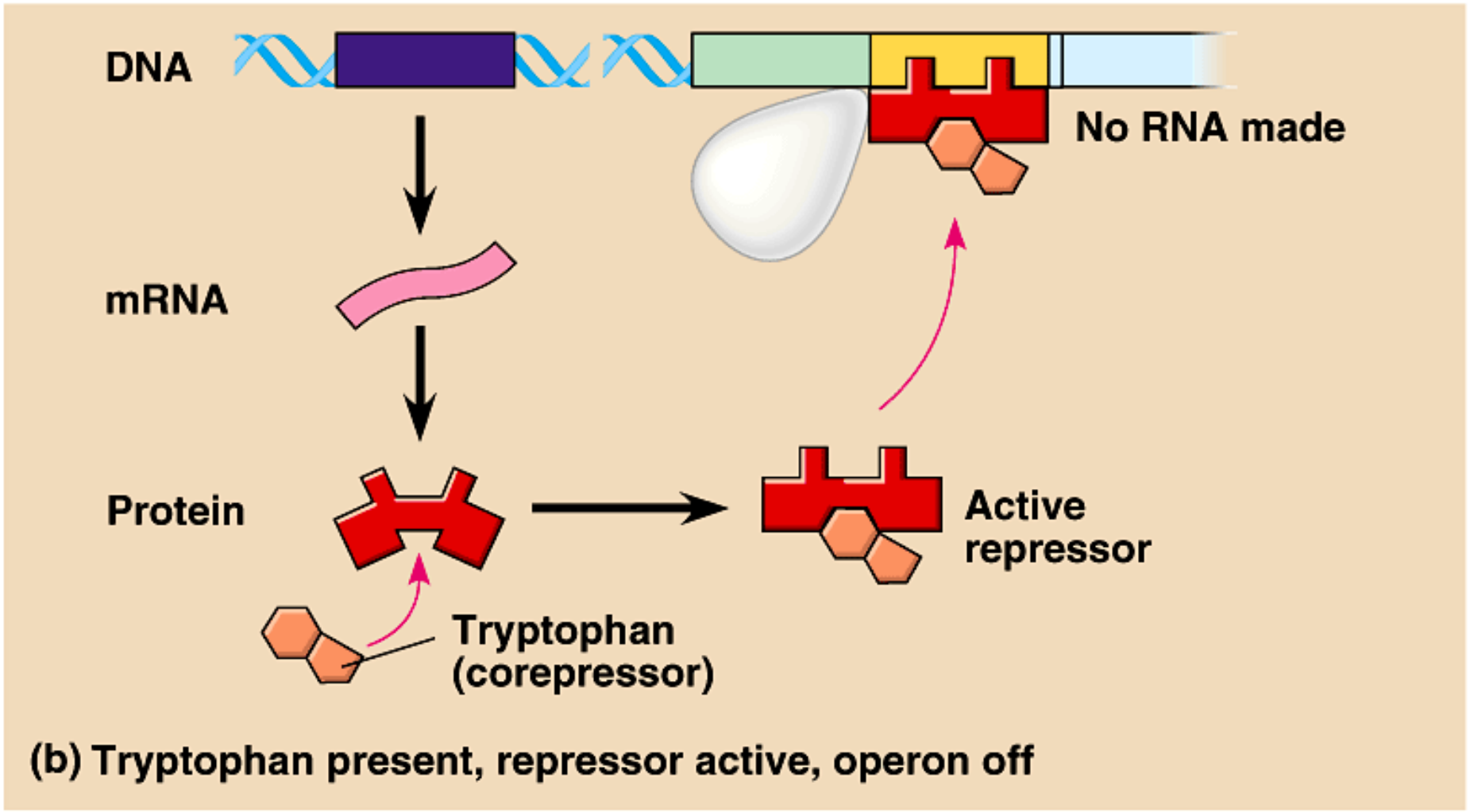
Repressor Bound to an Operator Blocks Transcription



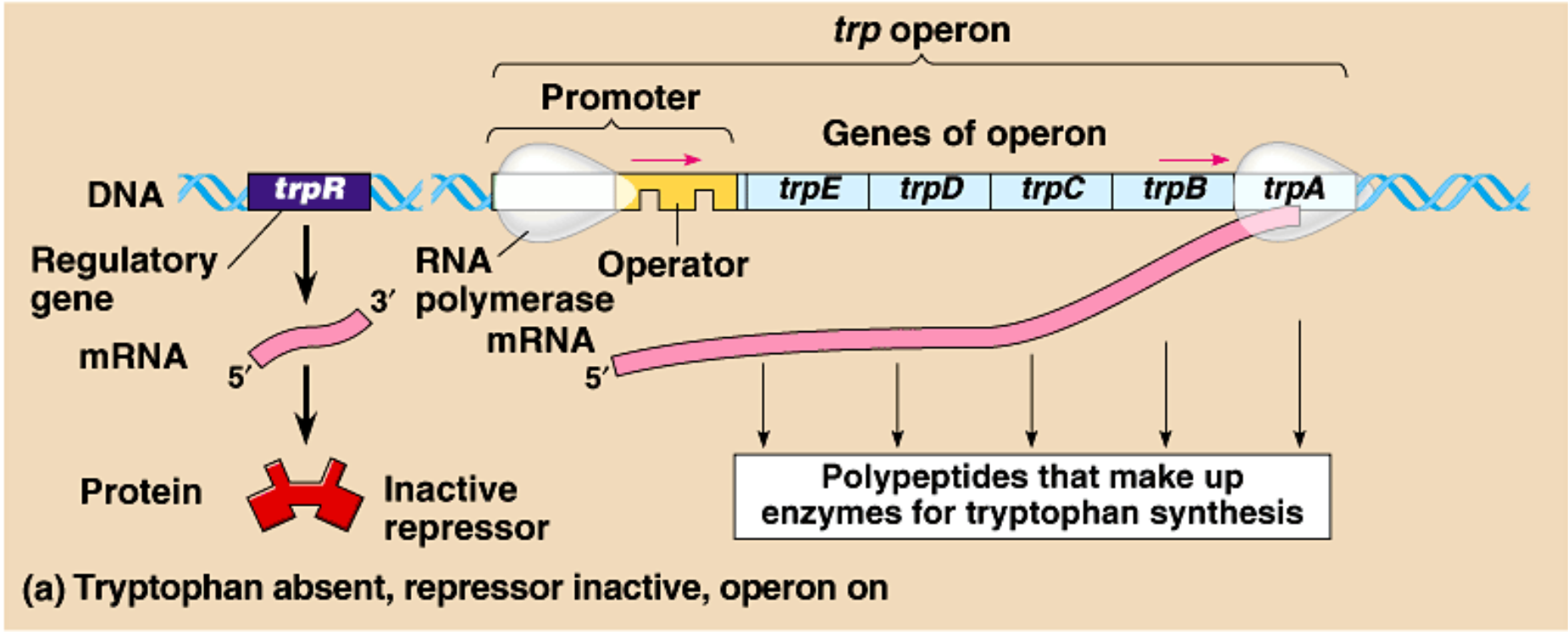
G. Regulation of Gene Expression in Prokaryotes

- The expression of prokaryotic genes is regulated by: inducible operator-repressor systems, repressible operator-repressor systems (e.g., both negative control), and systems that increase the efficiency of a promoter (e.g., positive control).
- Repressor proteins are coded by constitutive regulatory genes.

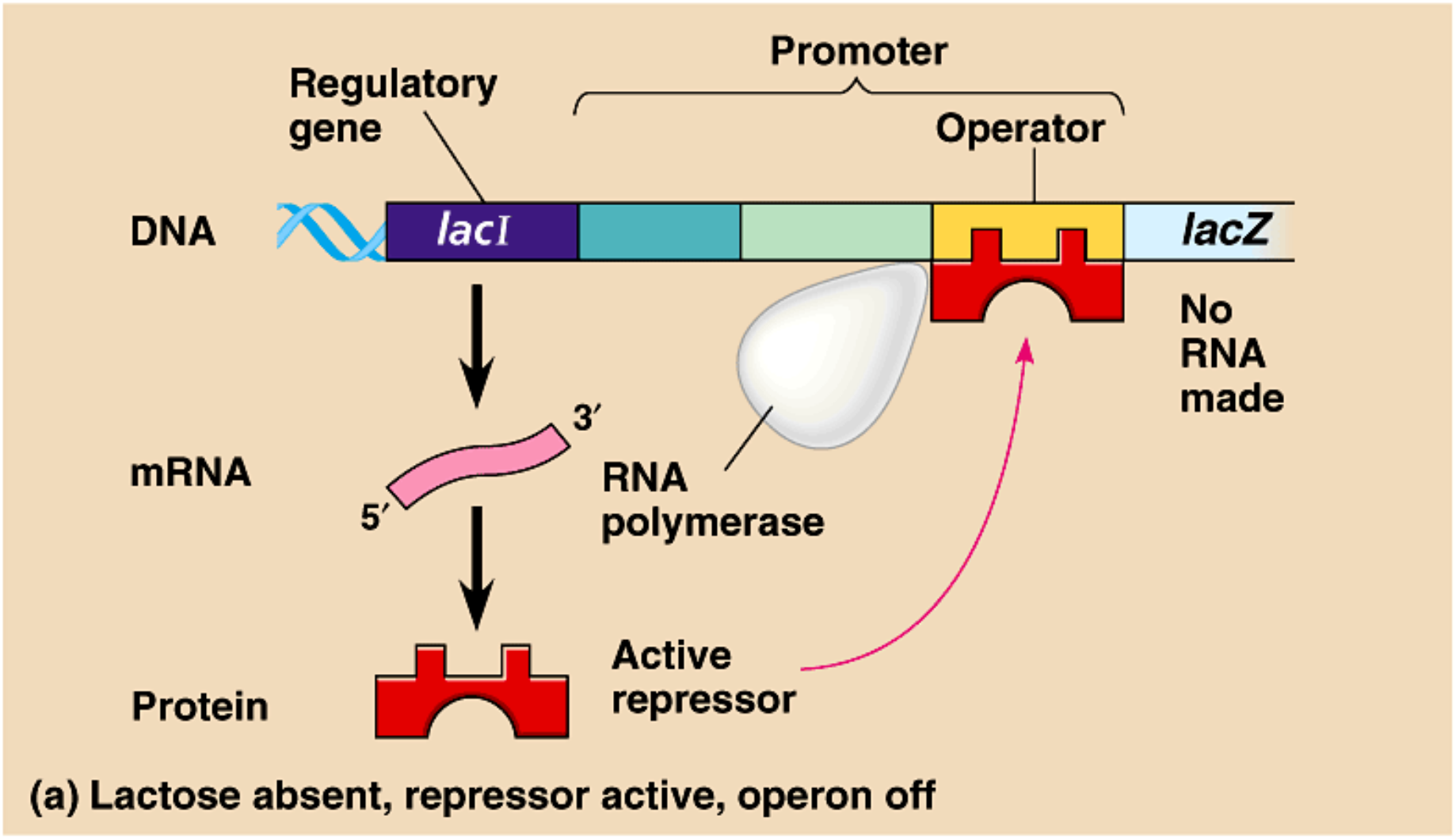
The *trp* operon: regulated synthesis of repressible enzymes



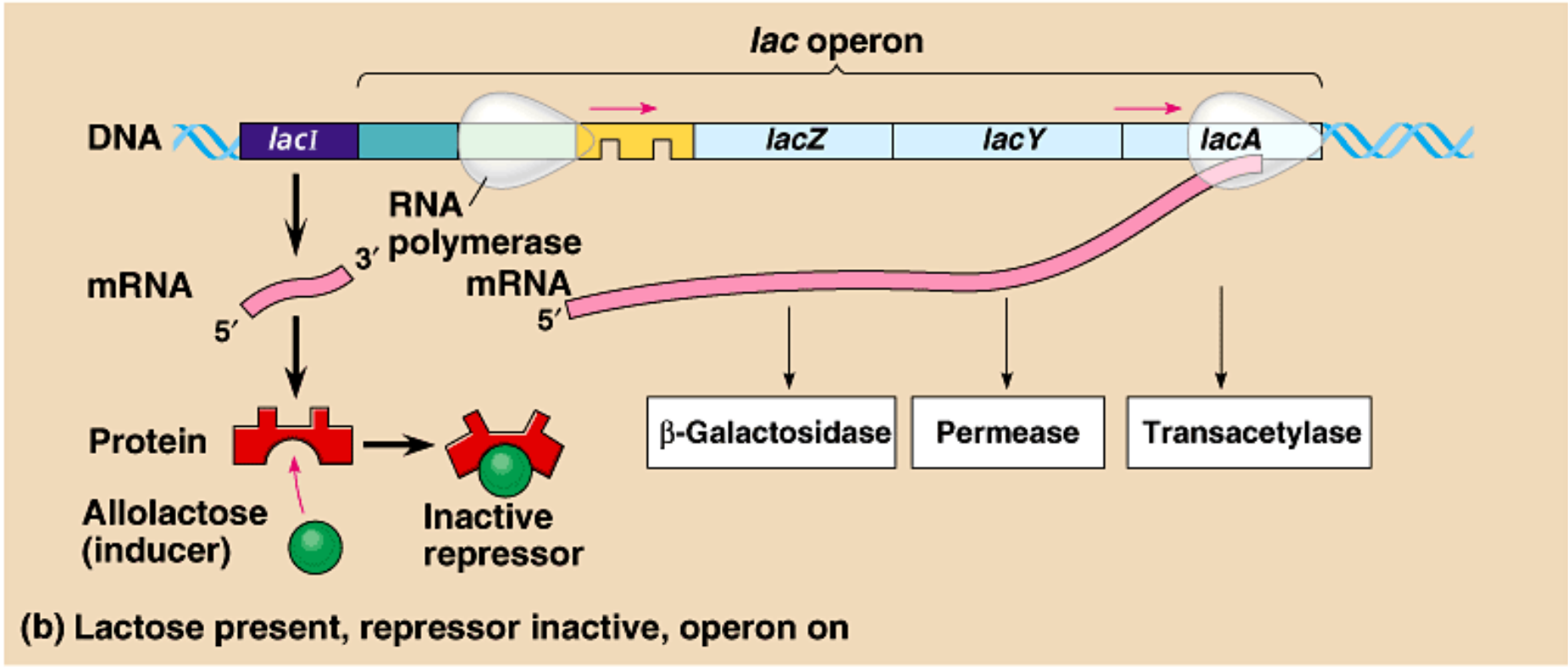
The *trp* operon: regulated synthesis of repressible enzymes



The *lac* operon: regulated synthesis of inducible enzymes



The *lac* operon: regulated synthesis of inducible enzymes

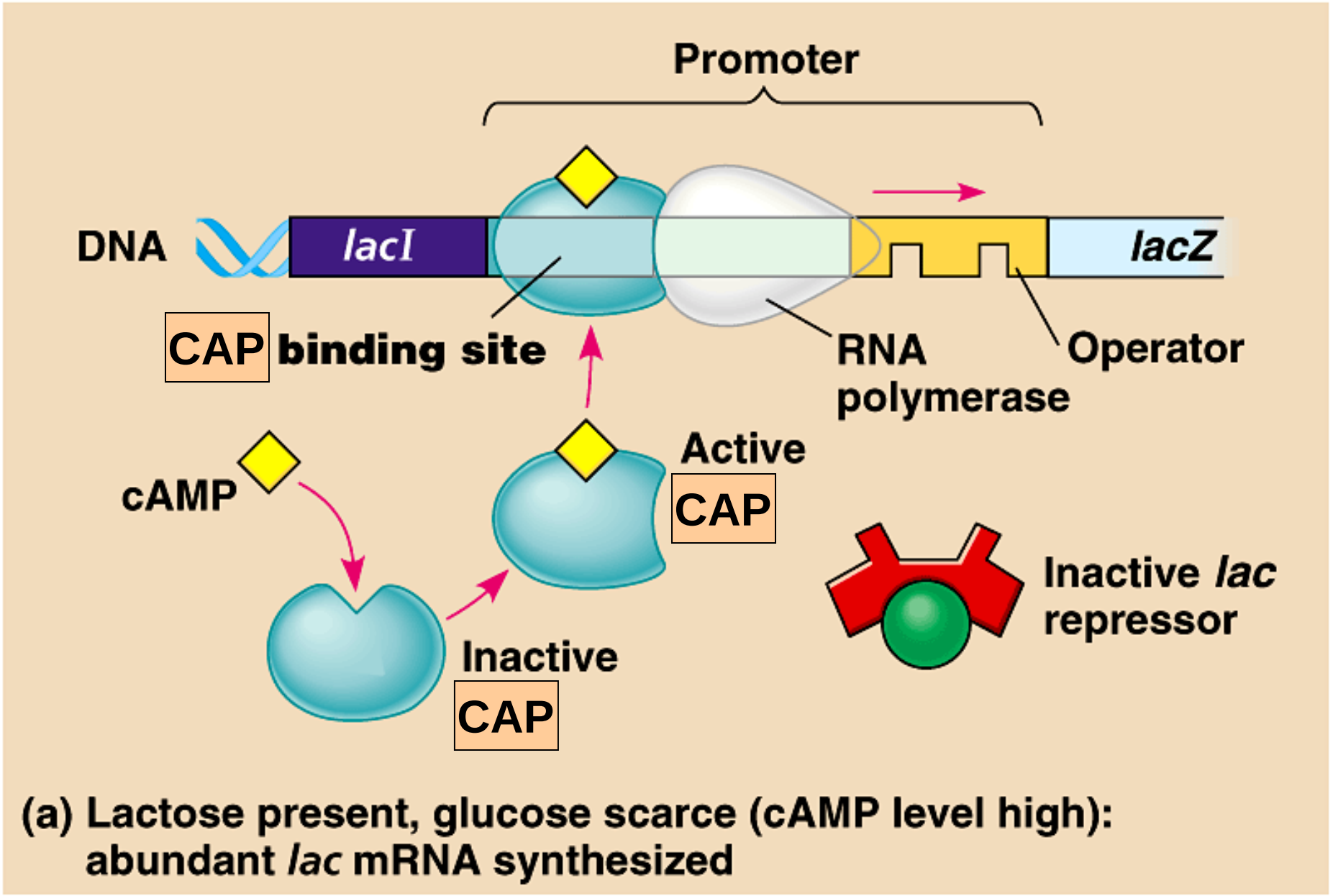


(b) Lactose present, repressor inactive, operon on

G. Regulation of Gene Expression in Prokaryotes

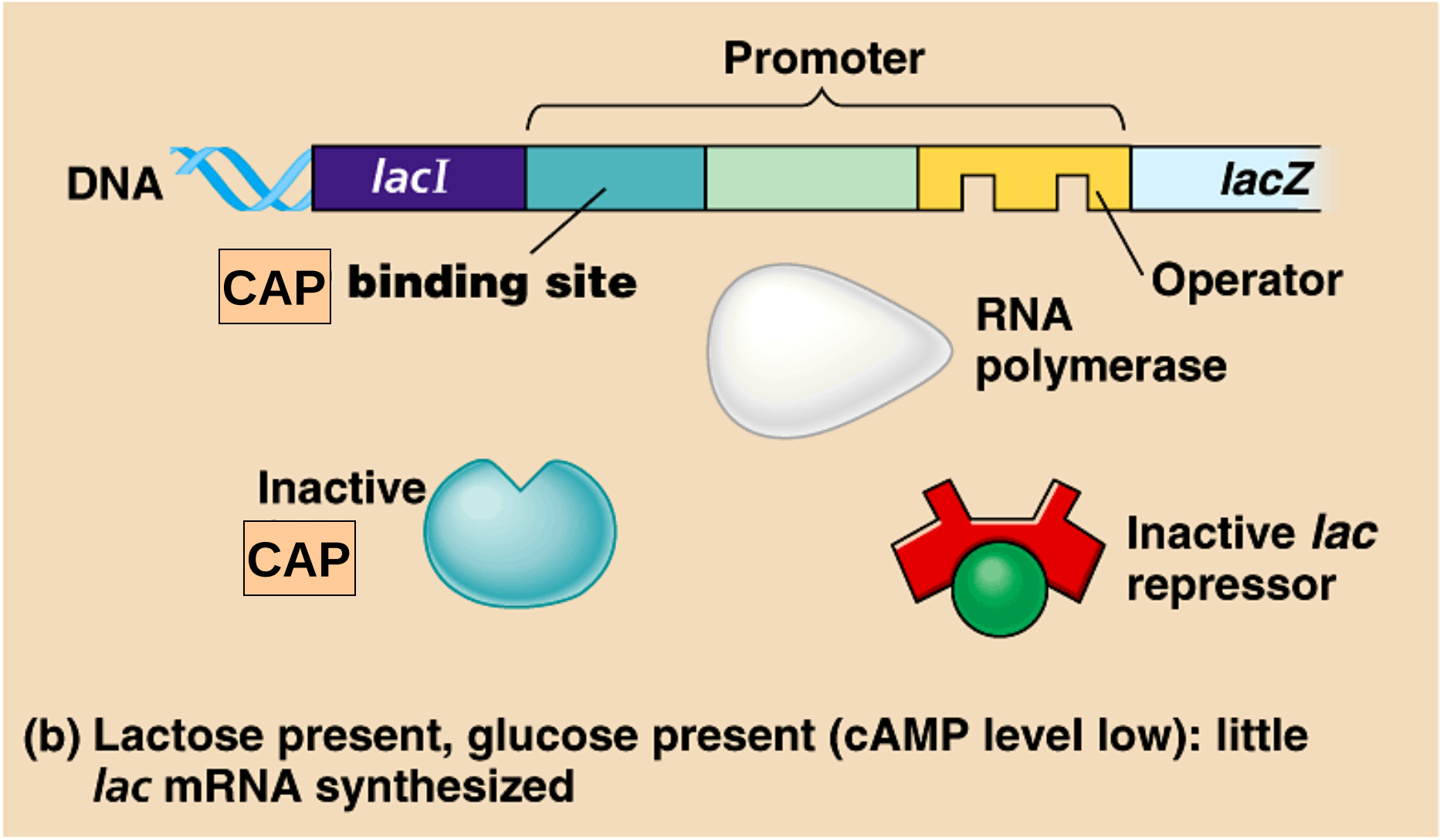
- The efficiency of RNA polymerase can be increased by regulation of the level of cyclic AMP, which binds to CAP (cAMP activator protein).
- The CAP-cAMP complex then binds to a site near the promoter of a target gene, enhancing the binding of RNA polymerase and hence transcription.

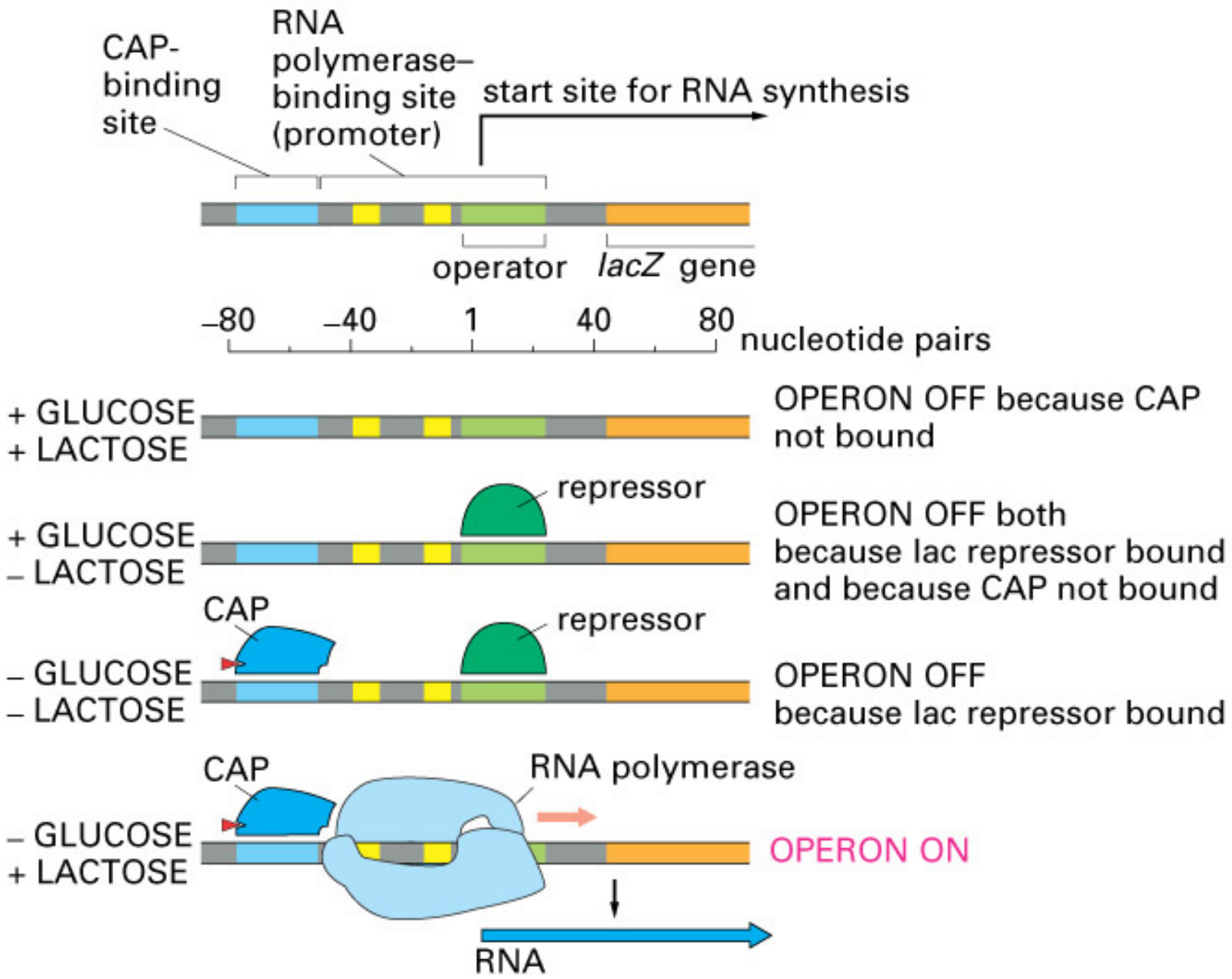
Positive control: cAMP activator protein



(a) Lactose present, glucose scarce (cAMP level high): abundant *lac* mRNA synthesized

Positive control: cAMP activator protein




CAP-binding site
RNA polymerase–binding site (promoter)
start site for RNA synthesis
operator
lacZ gene
–80
–40
1
40
80
nucleotide pairs
+ GLUCOSE
+ LACTOSE
OPERON OFF because CAP not bound
repressor
+ GLUCOSE
– LACTOSE
OPERON OFF both because lac repressor bound and because CAP not bound
CAP
repressor
– GLUCOSE
– LACTOSE
OPERON OFF because lac repressor bound
CAP
RNA polymerase
– GLUCOSE
+ LACTOSE
OPERON ON
RNA

13.2 The Relationships Between Positive and Negative Control in the lac Operon

GLUCOSE	cAMP LEVELS	RNA POLYMERASE BINDING TO PROMOTER	LACTOSE	*LAC* REPRESSOR	TRANSCRIPTION OF *LAC* GENES?	LACTOSE USED BY CELLS?
Present	Low	Absent	Absent	Active and bound to operator	No	No
Present	Low	Absent	Present	Inactive and not bound to operator	No	No
Absent	High	Present	Present	Inactive and not bound to operator	Yes	Yes
Absent	High	Absent	Absent	Active and bound to operator	No	No

H. Comparison of Control Features in Bacteria & Eucarya

- Bacteria have multiple genes under single control: operons
- Eucarya have multiple RNA polymerases
- Simple vs. Complex Transcription Factors
- Local vs. Distal Control: Enhancers/Silencers
- Eucarya must contend with Chromatin