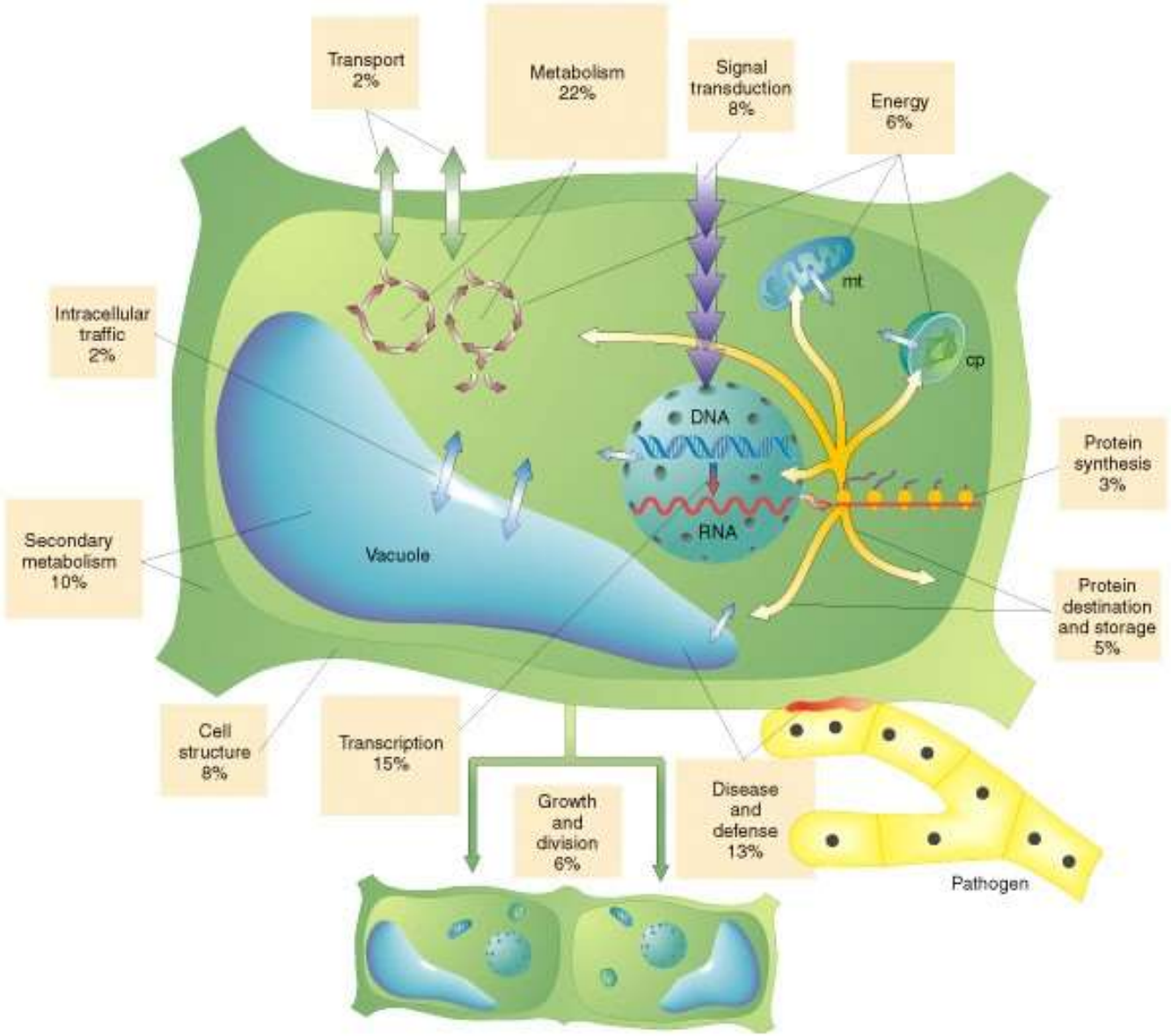
Peter Pristas

BNK1

Gene regulation in eukaryotes

Gene Expression in Eukaryotes



Only about 3-5% of all the genes in a human cell are expressed at any given time. The genes expressed can be specific for a particular cell type or tissue.

Galactose metabolism in yeast

- Galactose uptake and metabolism in yeast is encoded by the GAL genes.
- Like bacteria, a type of catabolite repression operates in yeast. Thus, galactose is utilized by the cell only if glucose is not present.
- The GAL genes are not physically linked and there no operons in yeast. However, expression of the GAL genes is coordinately regulated.

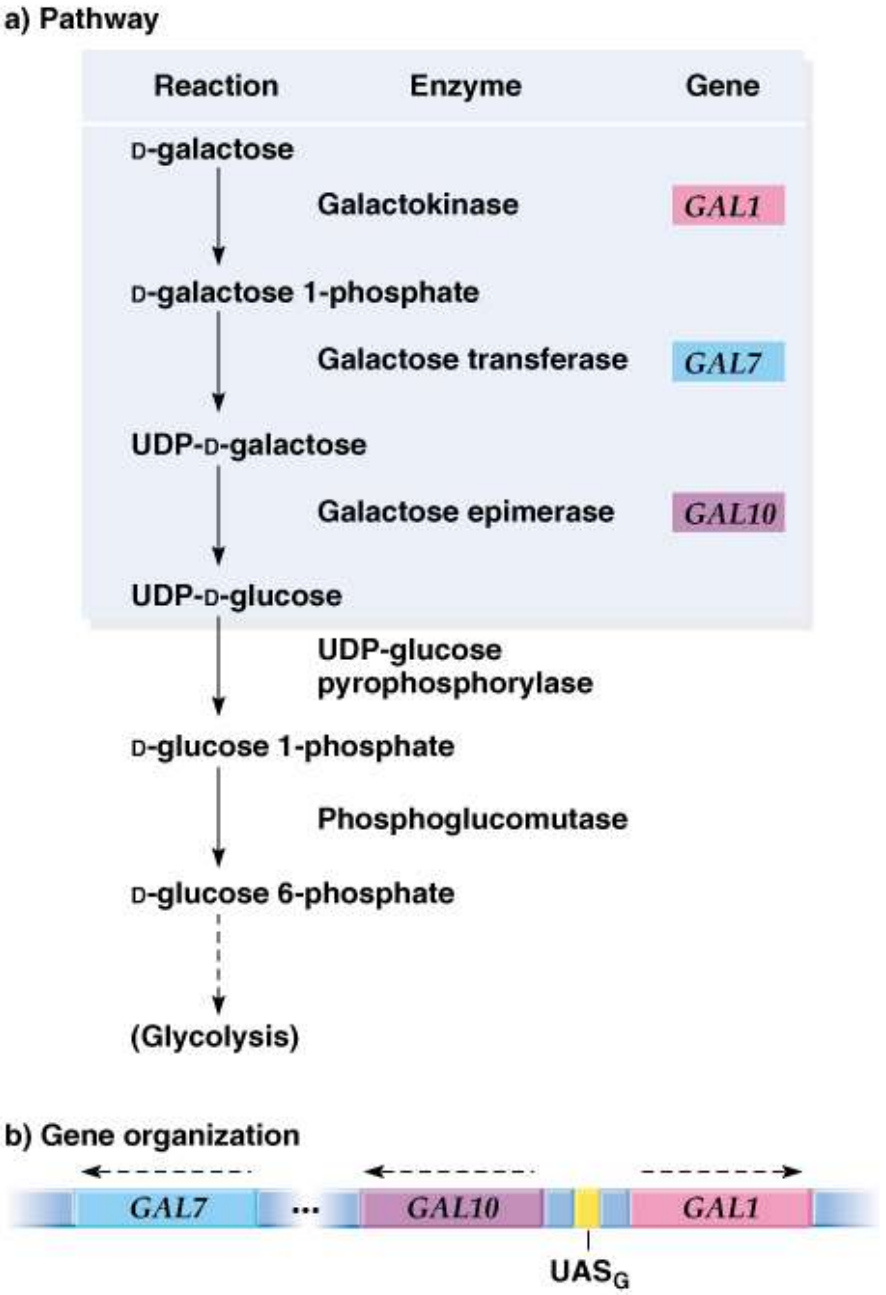
Galactose metabolism in yeast

TABLE 28–3 Genes of Galactose Metabolism in Yeast

	Protein function	Chromosomal location	Protein size (number of residues)	Relative protein expression in different carbon sources		
				Glucose	Glycerol	Galactose
Regulated genes						
GAL1	Galactokinase	II	528	−	−	+++
GAL2	Galactose permease	XII	574	−	−	+++
PGM2	Phosphoglucomutase	XIII	569	+	+	++
GAL7	Galactose 1-phosphate uridylyltransferase	II	365	−	−	+++
GAL10	UDP-glucose 4-epimerase	II	699	−	−	+++
MEL1	α-Galactosidase	II	453	−	+	++
Regulatory genes						
GAL3	Inducer	IV	520	−	+	++
GAL4	Transcriptional activator	XVI	881	+/−	+	+
GAL80	Transcriptional inhibitor	XIII	435	+	+	++

Source: Adapted from Reece, R. & Platt, A. (1997) Signaling activation and repression of RNA polymerase II transcription in yeast. *Bioessays* 19, 1001–1010.

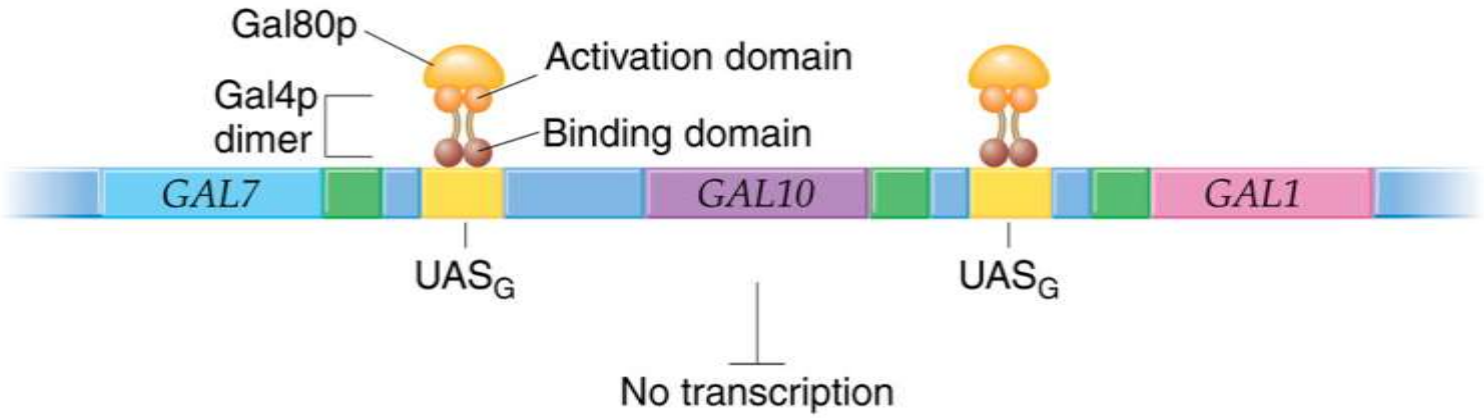
Galactose metabolism in yeast



Galactose metabolism in yeast

b) Absence of galactose

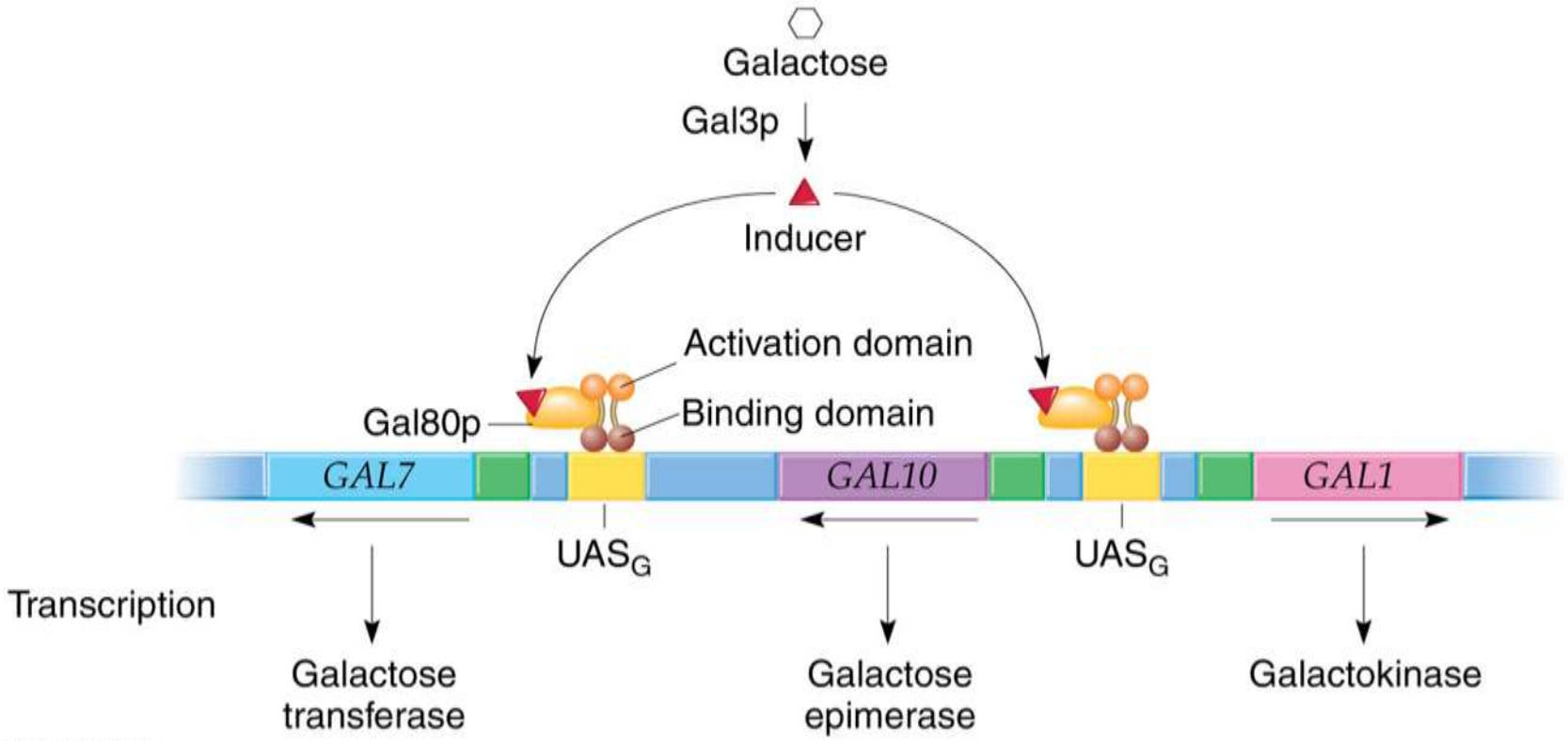
Gal80p binds to Gal4p activation domain, blocking it from activating transcription.



Galactose metabolism in yeast

c) Presence of galactose

Gal3p converts galactose to the inducer which binds to Gal80p, causing it to move on Gal4p. The now exposed Gal4p activation domain activates transcription.



GAL transcription

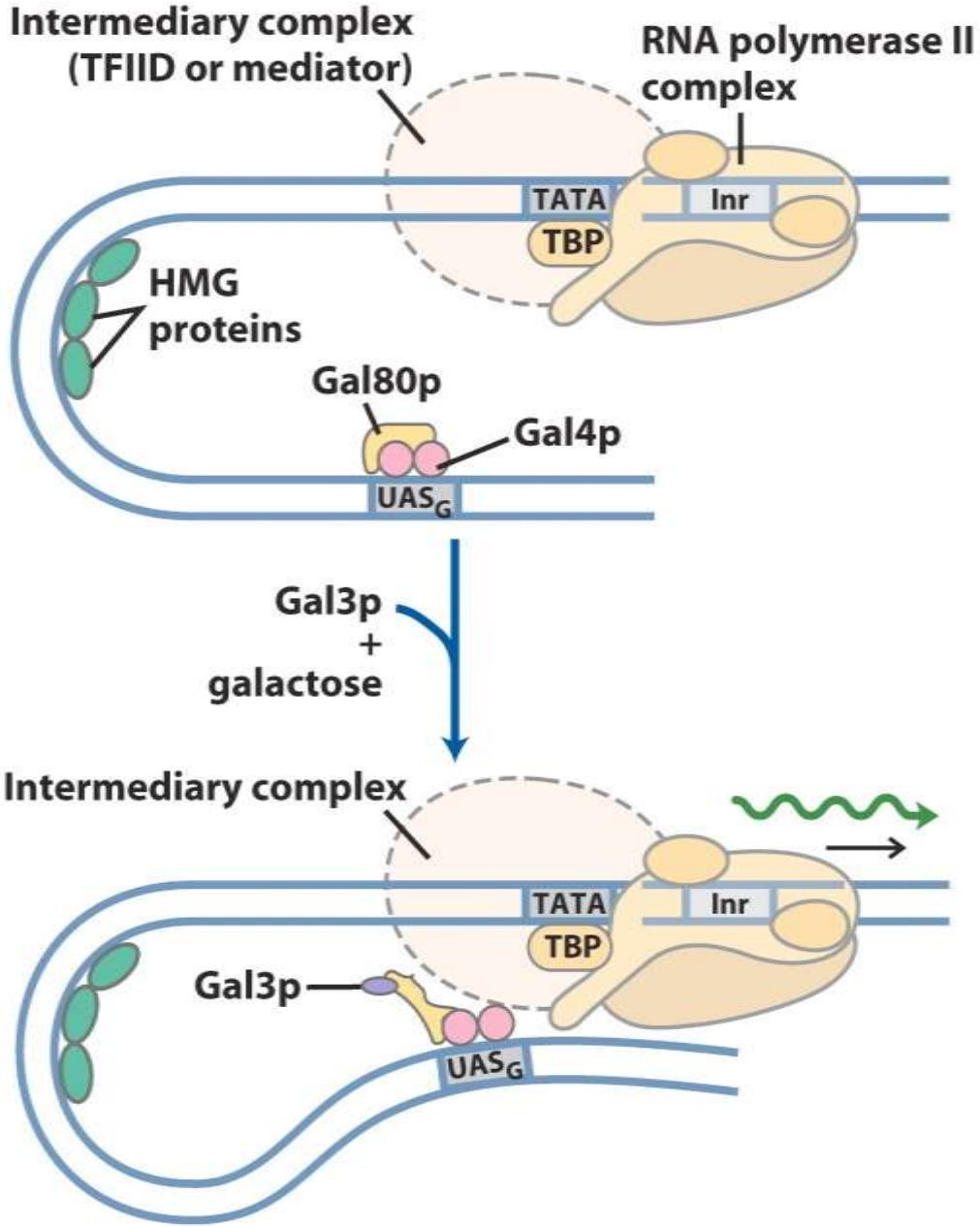
Gal4p is a transactivator of GAL gene transcription.

In the absence of galactose, Gal80p binds Gal4p and prevents its binding to UAS_G.

Galactose binds Gal3p, which binds Gal80p and exposes the DNA binding site on Gal4p.

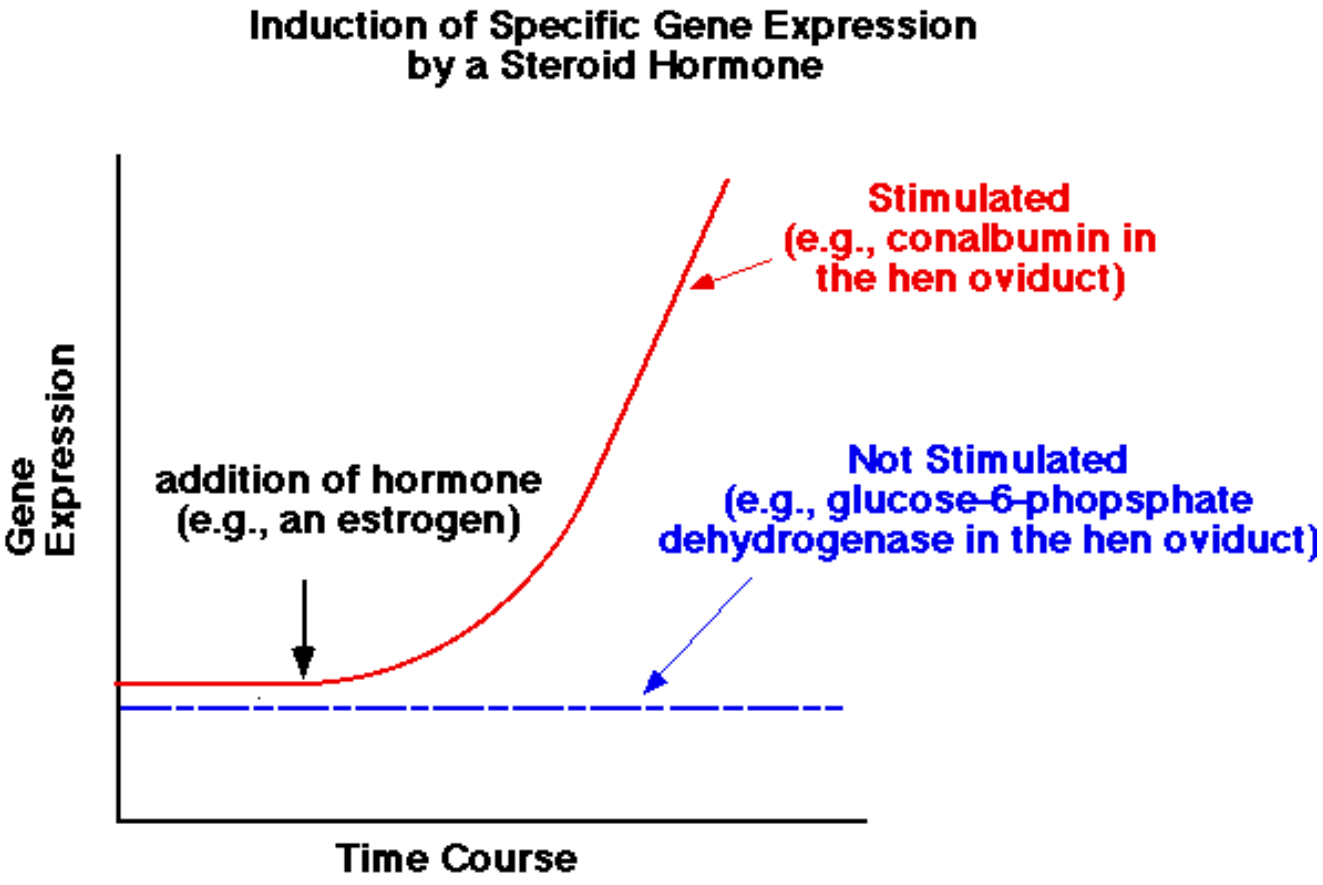
Transcription initiation occurs at GAL promoters.

Catabolite repression system is not shown.



Gene Expression in Eukaryotes

The control of eukaryotic transcription was originally thought to be identical to that observed in prokaryotes. "what's true for E. coli will be true for an elephant"



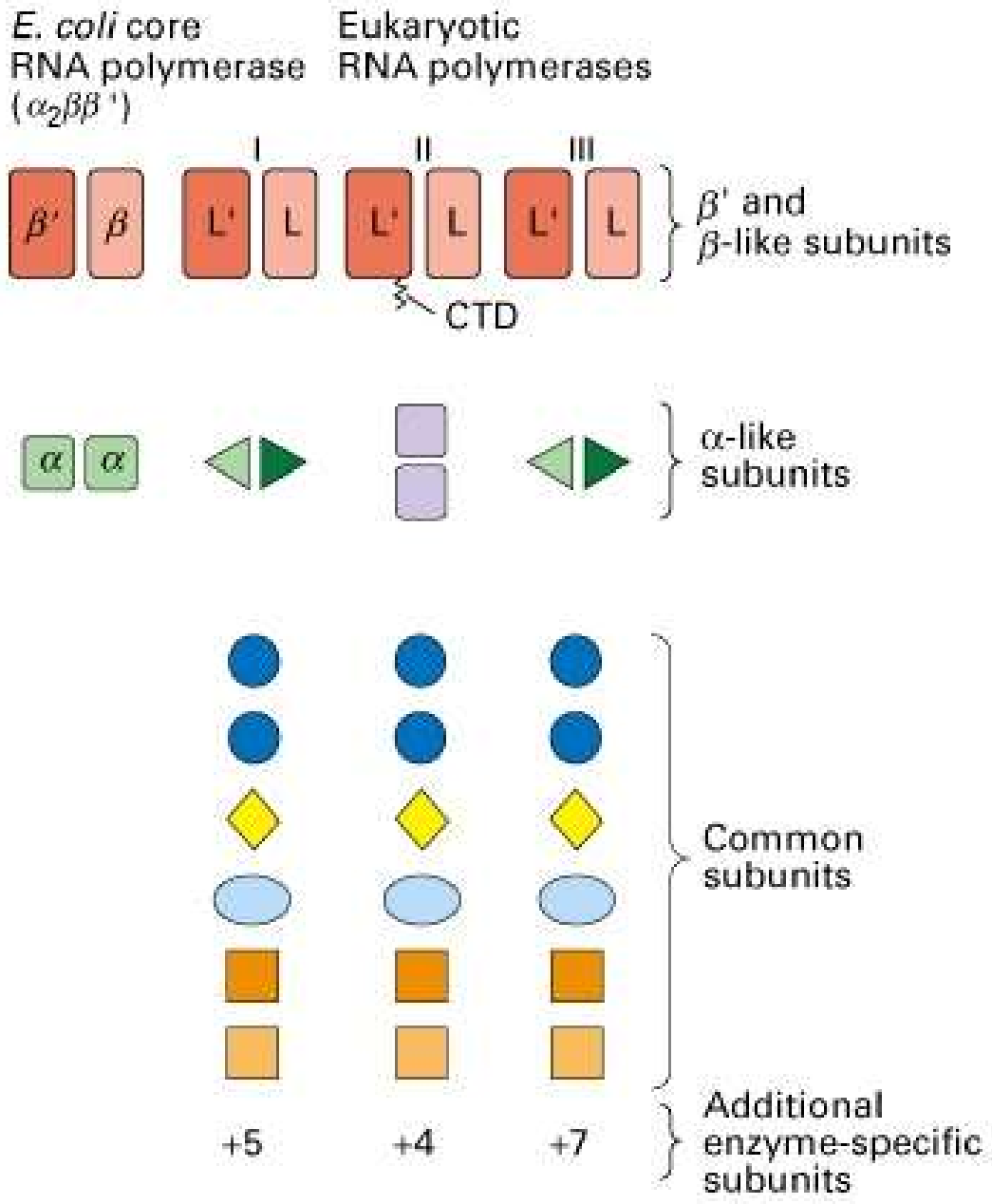
Transcription in eukaryotes

- **Generally the same as in prokaryotes except that**
  - more RNA polymerases
  - activity of transcription complex requires more accessory proteins.

- **Eukaryotic RNA polymerases**
  - 3 different RNA polymerases each transcribe different classes of nuclear genes:
    - RNA Pol I - transcribes large rRNA genes
    - RNA Pol II - transcribes protein encoding genes
    - RNA Pol III - transcribe small RNA, such as tTNA and 5S rRNA
  - All these RNA polymerases are complex, multisubunit enzymes. They differ from each other in subunit composition.
  - The 2 largest subunits of eukaryotic RNA pols are similar to b and b' subunits of *E. coli.*

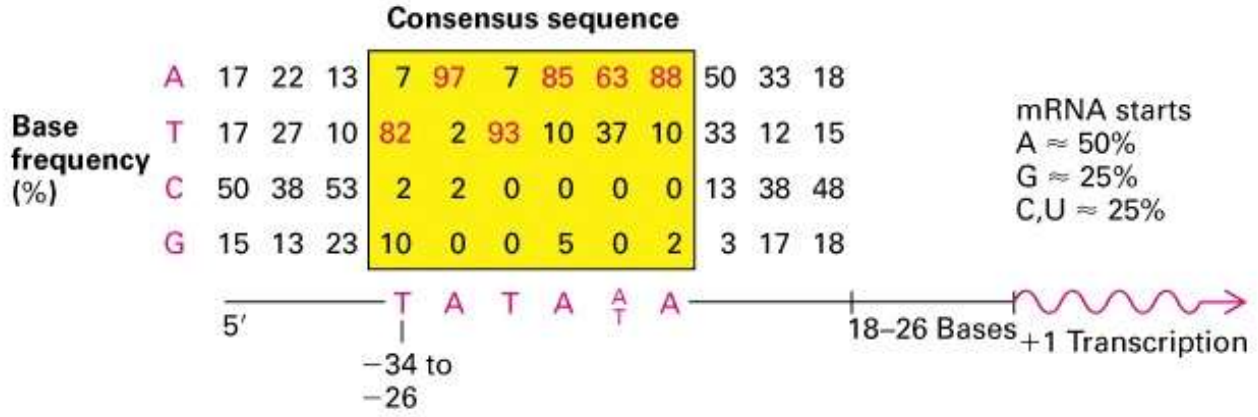
Regulation at transcription Level

Comparison of Prokaryotic and Eukaryotic RNA polymerases

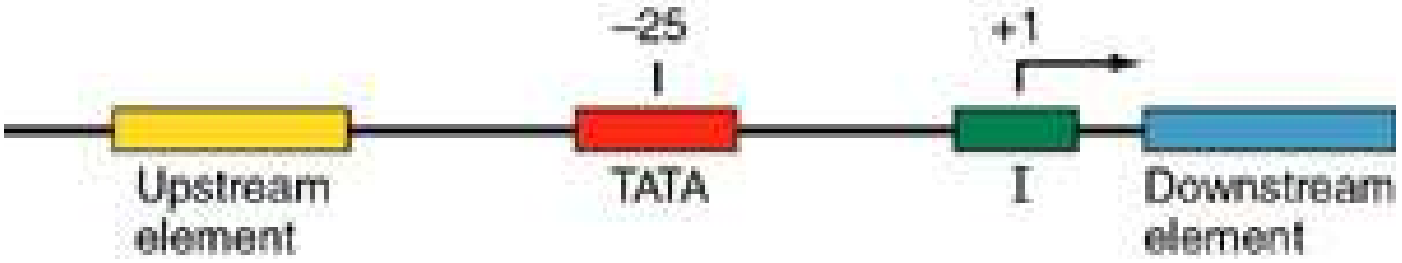


Promoters for RNA polymerase II Have a Very Limited Sequence Consensus and Occupy Larger Stretches of DNA

- **TATA box** (=Hogness box) promoters at -19 to -27 upstream from transcription start
- **Initiators**: C/A sequence consensus at –1/+1 sites
- **Upstream activating sequences - Sequences which control transcription initiation but do not necessarily confer the precision of transcription initiation, such as CpG islands**



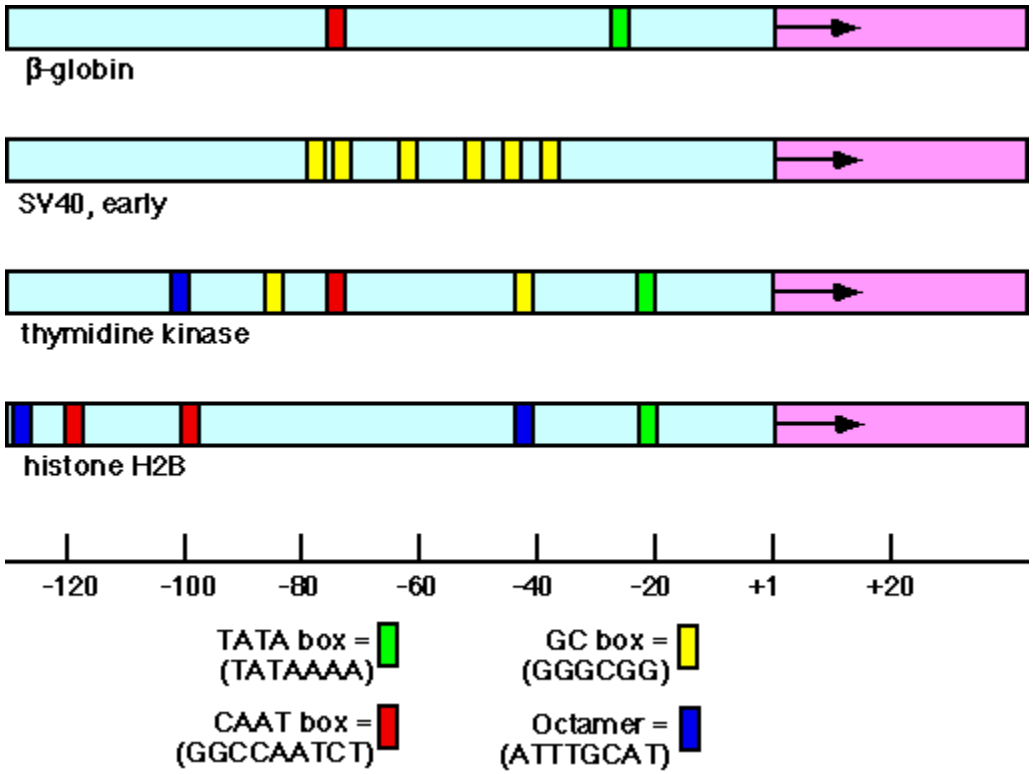
Eukaryotic Class II Promoters



The role of the TATA box appears to be the correct positioning of the polymerase to begin transcription.

Some genes do not have TATA boxes. Such genes, including the so-called "housekeeping" genes that are always expressed (constitutively "on") in a cell, have other elements that are involved in control, including the upstream elements.

Eukaryotic Class II Promoters



Eukaryotic Class II Promoters

Downstream Elements:
There is not consensus sequence here. These elements are only known by the fact that experimental alteration of them changes the transcription of the genes that contains them.

Enhancers and Silencers:
The rate of transcription in prokaryotes can be altered by binding of proteins to specific sequences (remember cAMP-CAP binding). In eukaryotic cells there are sequences that influence the rate of transcription by interaction with specific proteins.
Sequences that increase the rate of transcription are called enhancers. The can be located near the transcription unit in question or quite distant from it.

Eukaryotic transcription factors

- Assembly of transcription complex more complicated in eukaryotes
- RNA pol do not bind nonspecifically to promoters on their own. Need transcription factors (Tfs).
- Most TFs isolated not as part of transcription complex.
- Tfs either bind directly to DNA before RNA Pol binds, or associate with RNA Pol after it binds to promoter.
- ***General transcription factors*** = required for transcription of all genes of a class.
- Other transcription factors may be required only by certain genes or gene families.
- All eukaryotes contain a number of similar general TFs that interact directly with RNA Pol II and control initiation at all class II genes.
- Many class II genes have ***TATA box*** between 19 and 27 bp upstream of transcription start site. TATA box in eukaryoters is where RNA pol II binds DNA during assembly of initiation complex. (TATA box recognized by TBP, a subunit of TFIID)
- TF's bind promoter before RNA Pol II and attract RNA Pol II to promoter.

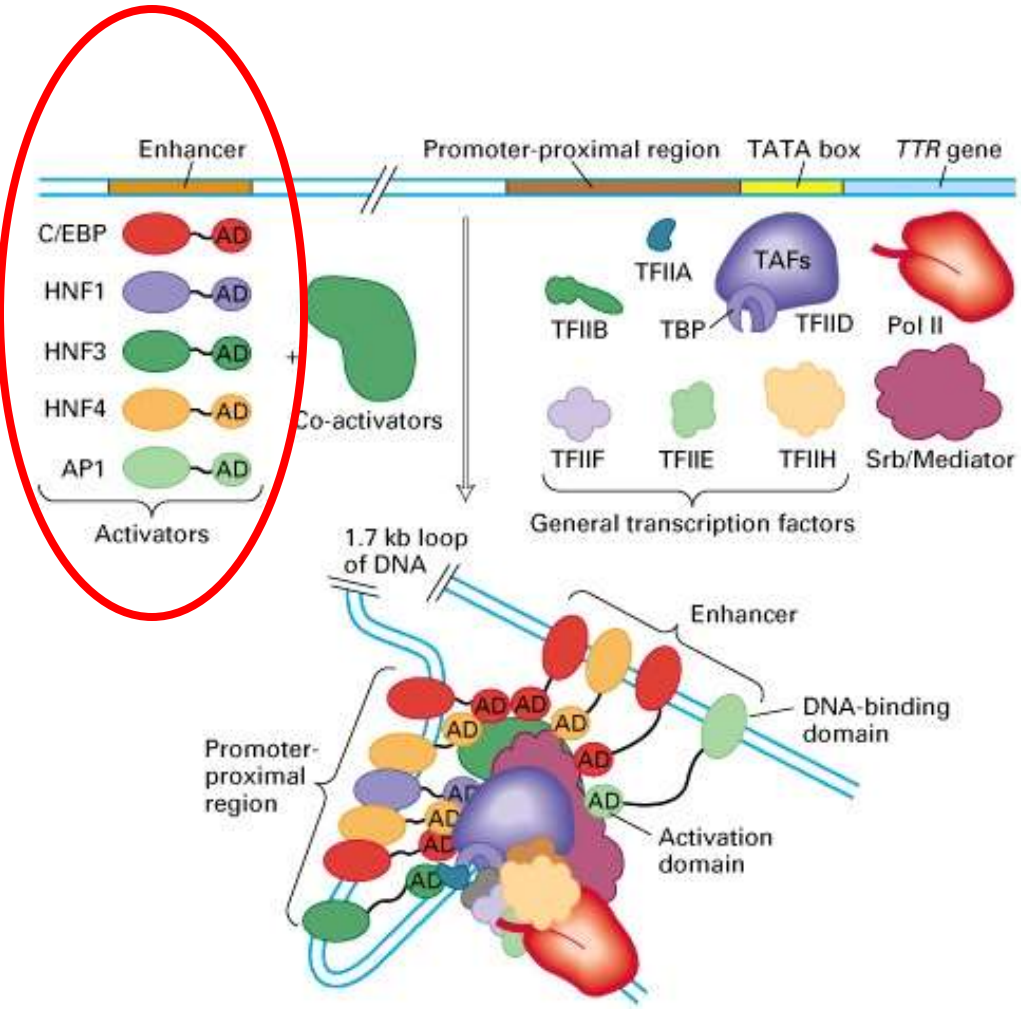
Eukaryotic transcription factors

- recognize target sequences in DNA
- interact with other transcription factors

These trans-acting factors can control gene expression in several ways:

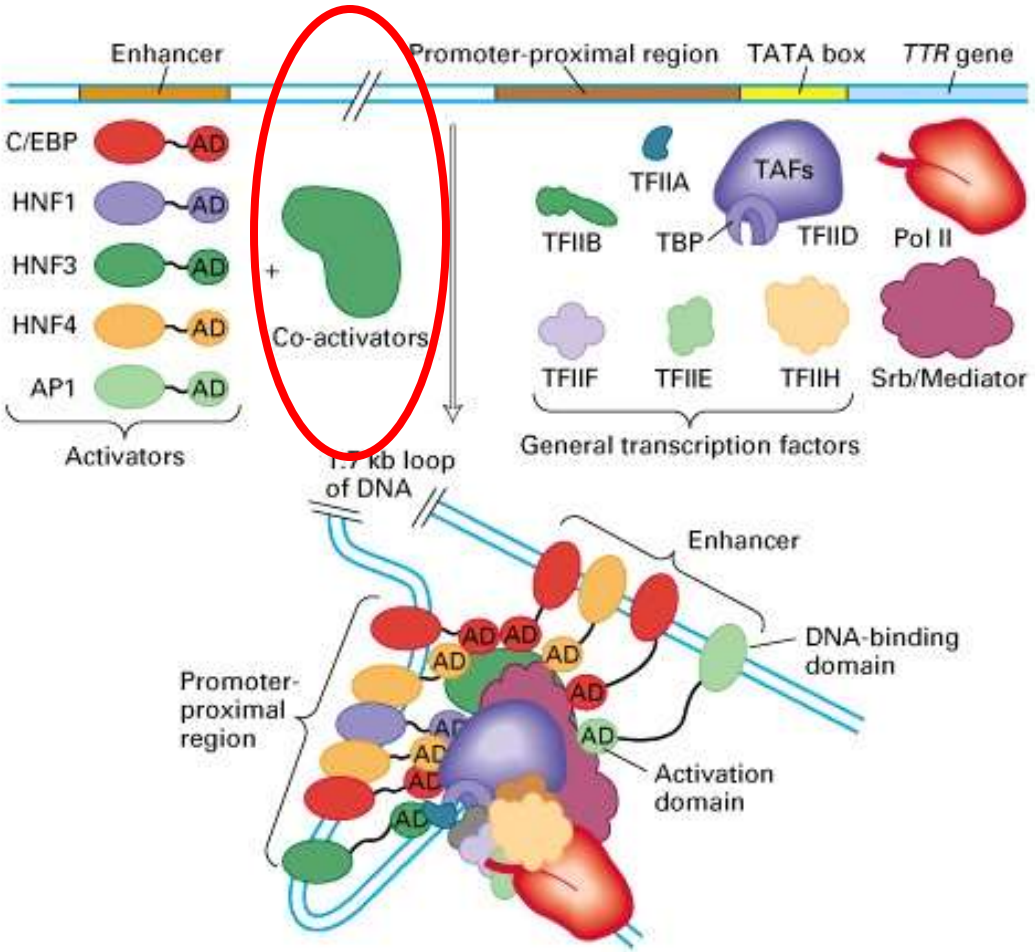
- factor may be expressed in a specific tissue manner (spatial regulation)
- factor may be expressed in at specific time in development (temporal regulation)
- factor may require modification (phosphorylation)
- factor may be activated by ligand binding
- factor may be sequestered until an appropriate environmental signal allows it to interact with the nuclear DNA

Enhancer binding proteins



- Also known as DNA binding transactivators
  - These bind enhancers that are far away from the promoter
    - They recognize the specific enhancer sequence
    - Some enhancer binding proteins work on a large number of genes, permitting coordinate control of transcription
    - Others are specific to a single gene
  - They then loop inward toward the promoter so that the enhancer binding protein can interact with the basal transcription factors at the promoter site
    - Protein-protein interactions are mediated through motifs such as the leucine zipper and the helix loop helix

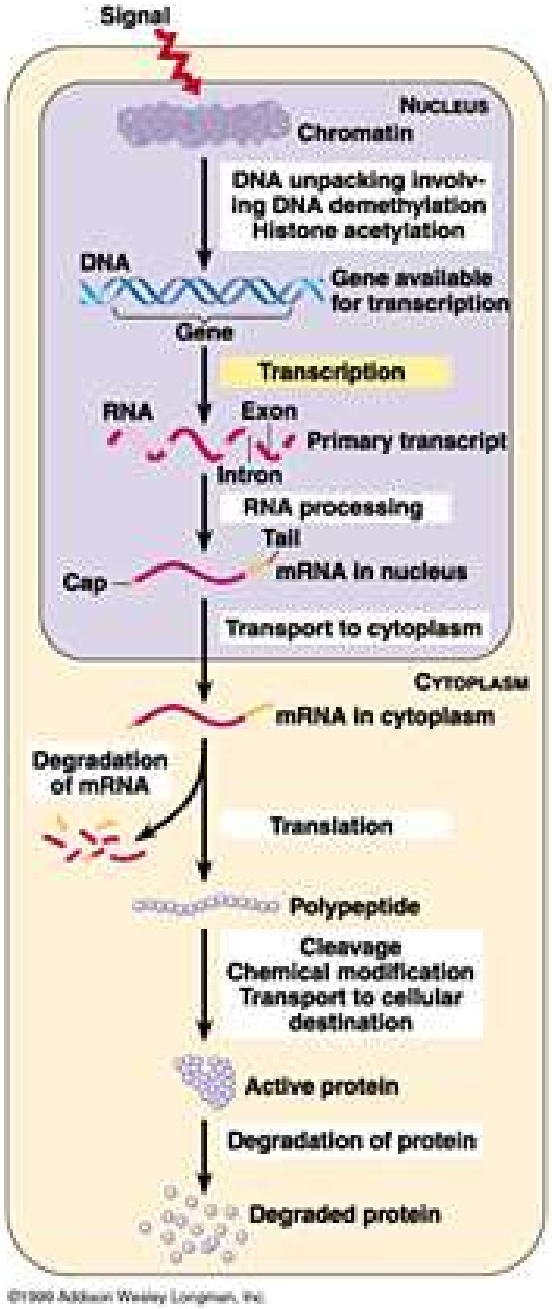
Coactivator proteins



- These bind RNA polymerase II complexes and enhancer binding proteins and mediate the signaling between them
- RNA polymerase II may carry the coactivator proteins with it as it transcribes
- Coactivators are necessary for transcription

In eukaryotes, one must consider

- **DNA is in chromatin**
- **There are three RNA polymerases versus one in prokaryotes**
- **The initial transcript requires processing before it is functional mRNA**
- **mRNA is synthesized in nucleus but translated in cytoplasm so transcription-translation coupling does not occur**



Eukaryotic gene regulation occurs at several levels

Given the spectrum of cell, tissue and organ types, a high degree of regulation must be available.

There is a means for regulation at every step from transcription to translation.

Transcription is confined to the nucleus

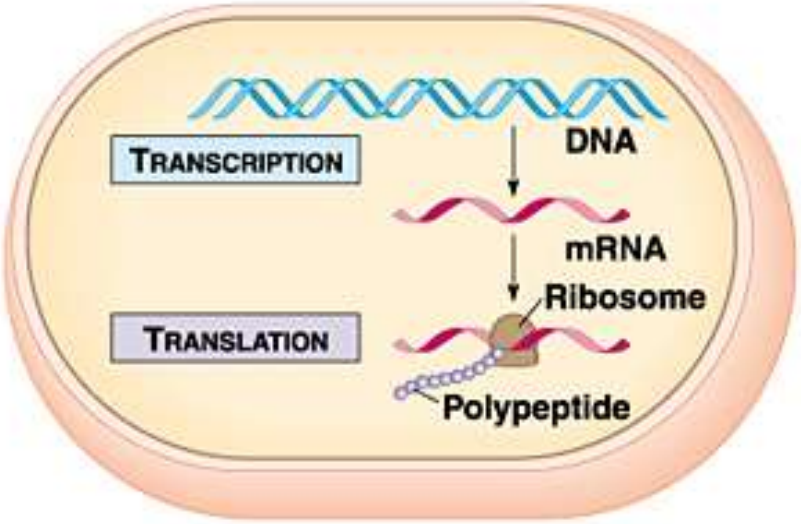
Translation is confined to the cytoplasm

most eukaryotic genes are regulated by multiple transcription control elements

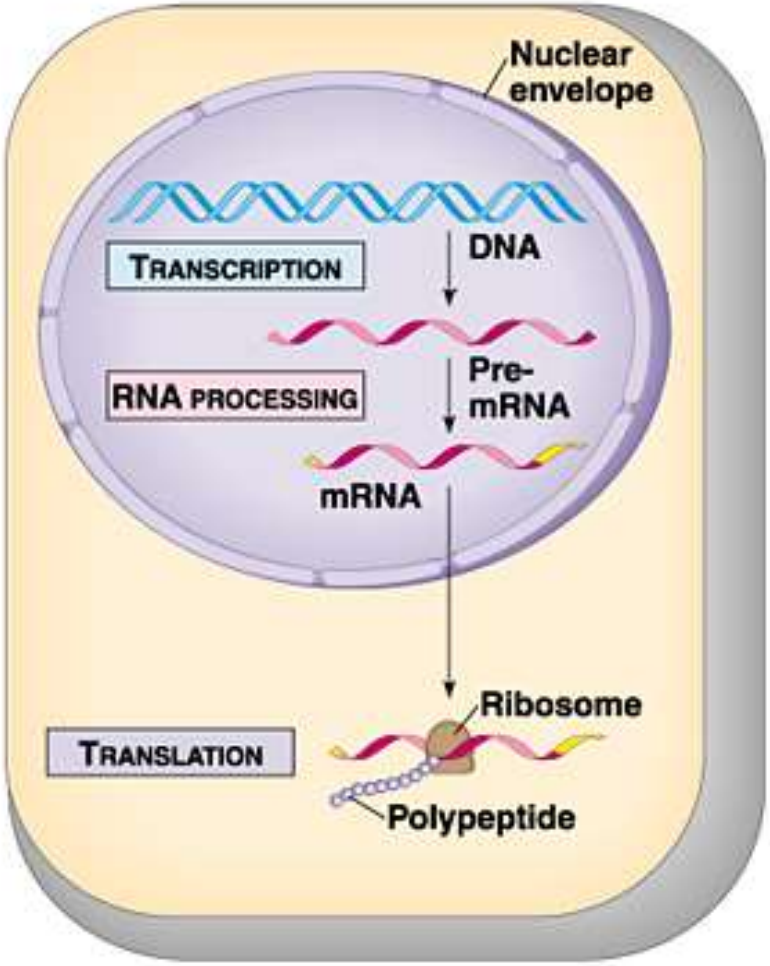
Typical half-lives of mRNA molecules

Cell	Generation time of cell	mRNA Half-Lives	
		Average	Range
Escherichia coli	20 - 60 min	3 - 5 min	2 - 10 min
Saccharomyces cerevisiae (yeast)	3 hr	22 min	4 - 40 min
Cultured human or rodent cells	16 - 24 hr	10 hrs	30 min or less (histone, *c-myc* mRNAs) 0.3 - 24 hr (specific mRNAs)

Prokaryotes and Eukaryotes



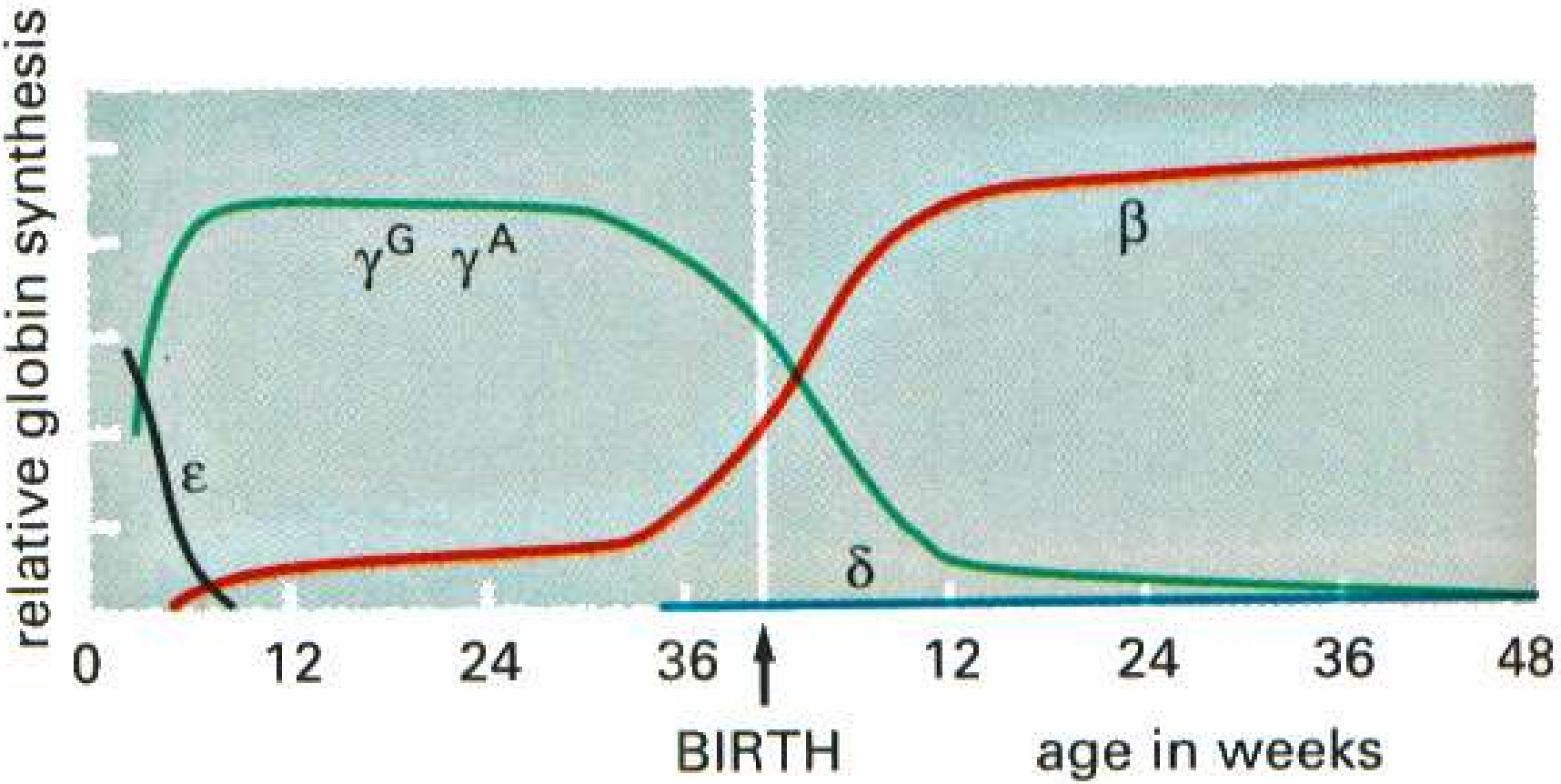
(a) Prokaryotic cell



(b) Eukaryotic cell

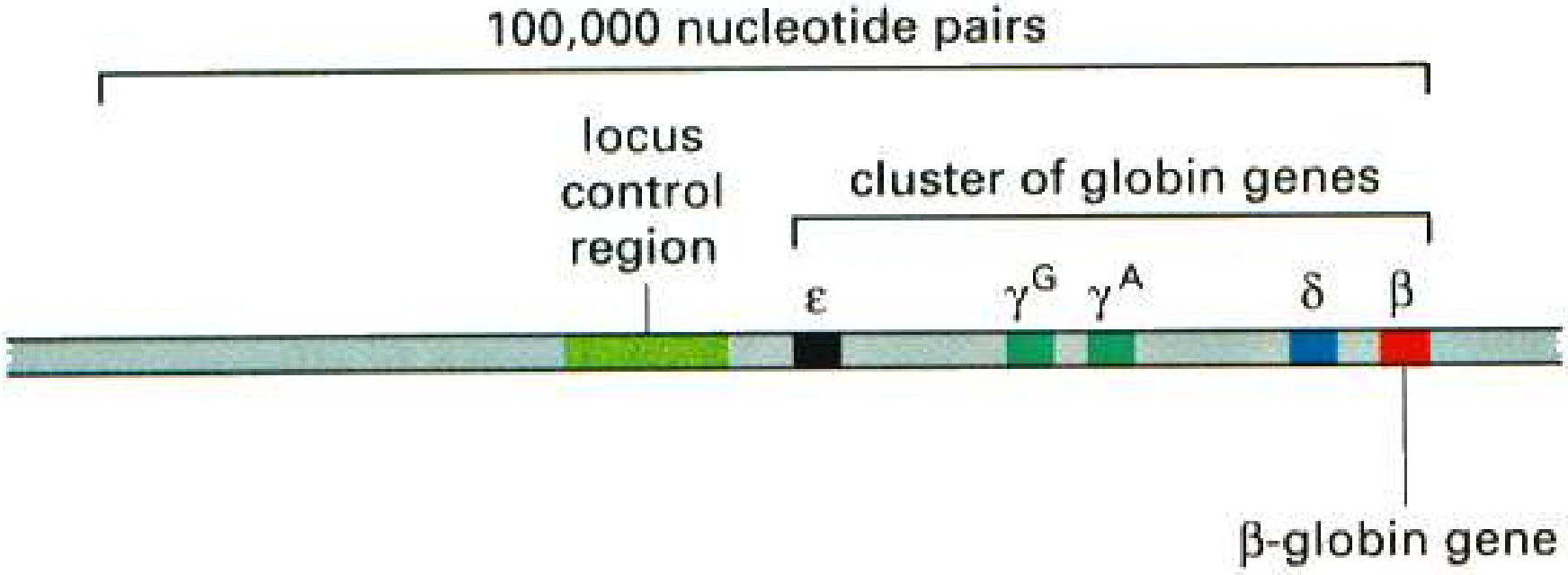
- prokaryotes (bacteria) do not have nuclei
- eukaryotes segrege transcription in the nucleus. mRNA is also preprocessed prior to translation in eukaryotes

The differential expression of beta-globin genes during human development

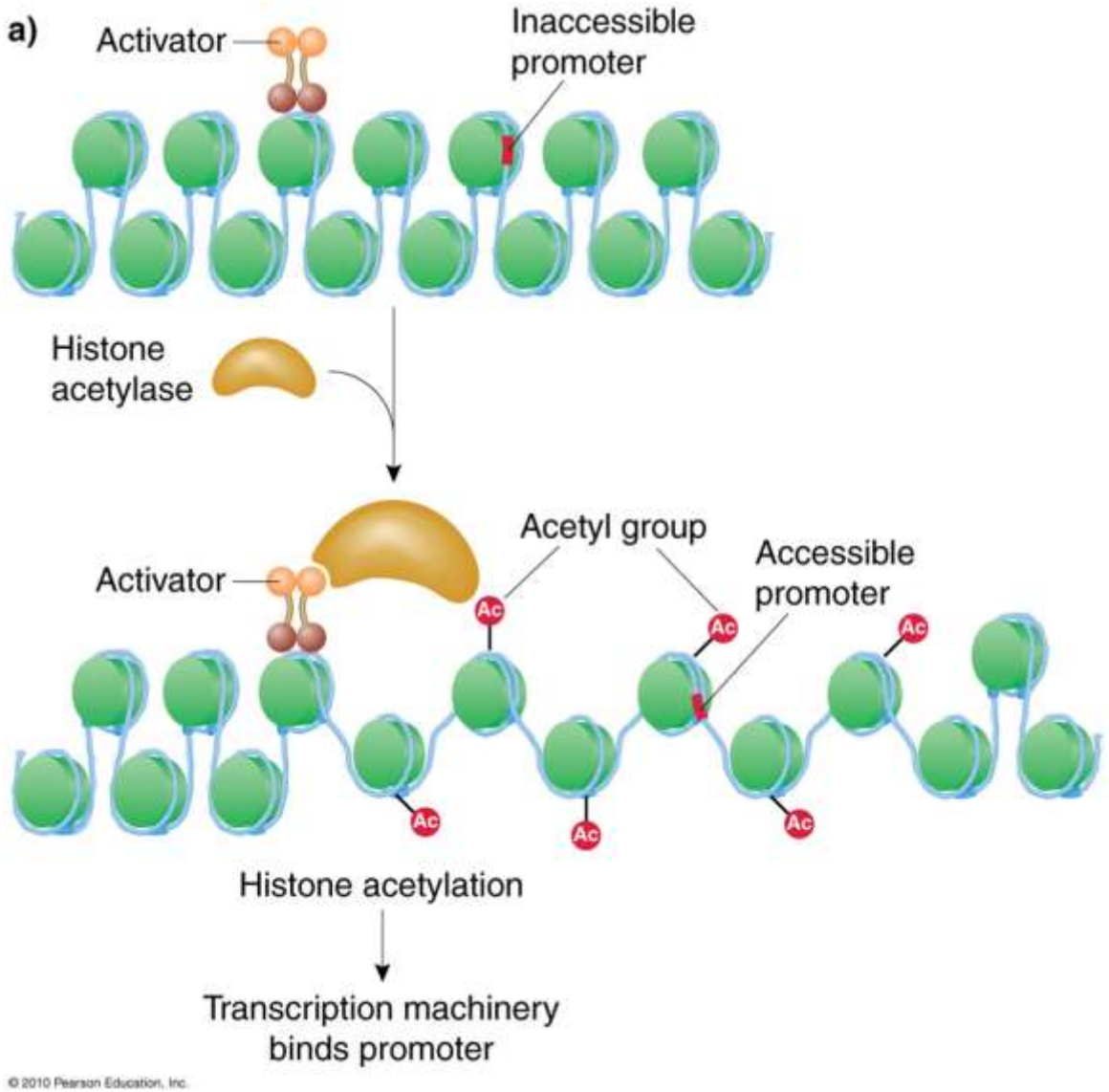


The cluster of beta-like globin genes in humans

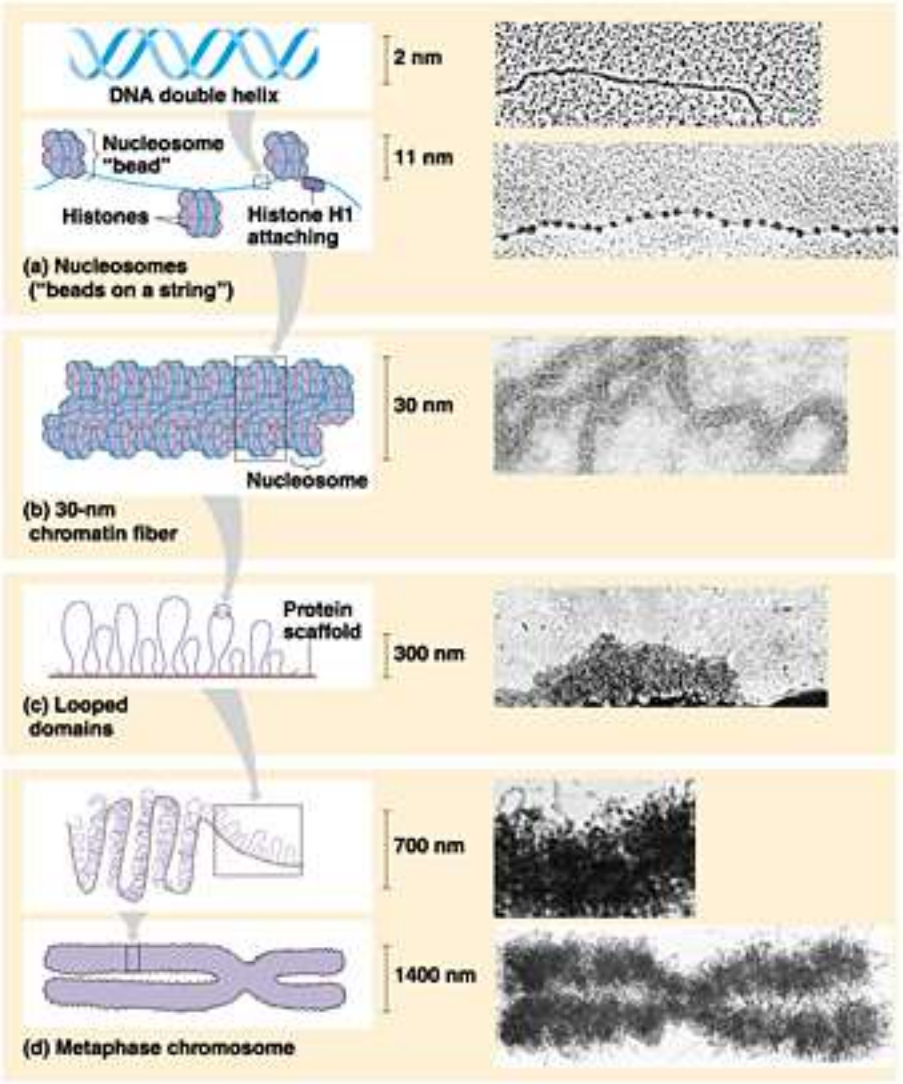
The region contains 100,000 nucleotide pairs and 5 globin genes and a locus control region. Each globin gene is regulated by a specific set of regulatory proteins . In addition, the entire cluster is subject to an on-off control that involves global changes in chromatin structure.



Two stages are involved in the regulation of the human globin gene cluster



Chromatin structure



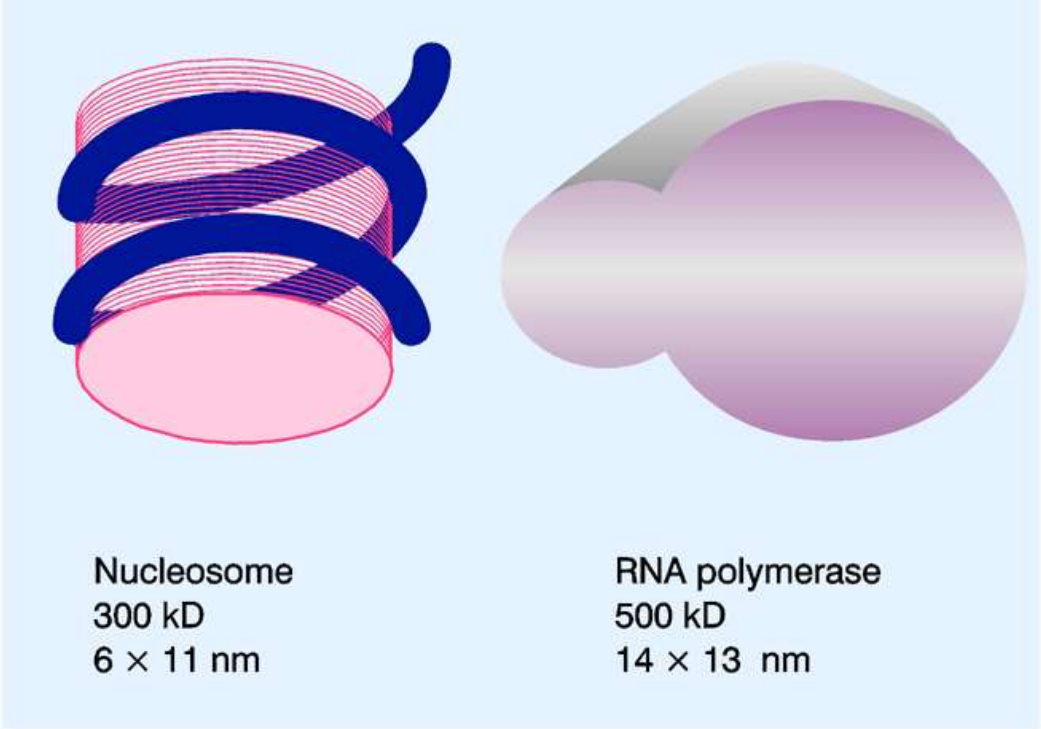
In order to pack a very long piece of DNA into the smallest space, the DNA is successively wound up around proteins and itself

Unwound human DNA is ~ 6 cm long !

One very important structure is the nucleosome. Genes can be transcribed when there are histones in the way. But this assembly is very dynamic (histones can be released and assembled)

Most of a prokaryotic genome encodes proteins. In contrast, 97% of the human genome appears to encode nothing !

Chromatin structure

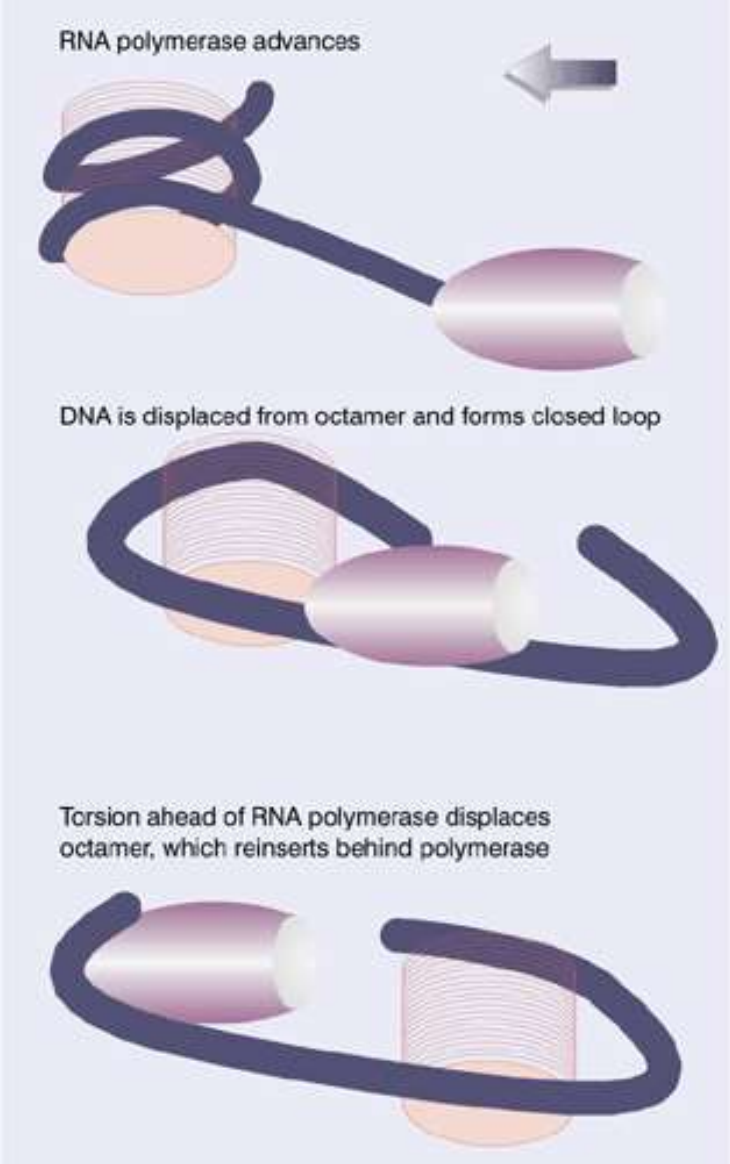


RNA polymerase is comparable in size to the nucleosome and might encounter difficulties in following the DNA around the histone octamer.

Some Enzyme Complexes Catalyzing Chromatin Structural Changes during Transcription

Enzyme complex*	Oligomeric structure	Source	Activities
GCN5-ADA2-ADA3	3 polypeptides	Yeast	GCN5 has type A HAT activity
SAGA/PCAF	>20 polypeptides	Eukaryotes	Includes GCN5-ADA2-ADA3
SWI/SNF	>11 polypeptides; M_r 2×10^6	Eukaryotes	ATP-dependent nucleosome remodeling
NURF	4 polypeptides; M_r 500,000	*Drosophila*	ATP-dependent nucleosome remodeling
CAFI	>2 polypeptides	Humans; *Drosophila*	Responsible for binding histones H3 and H4 to DNA
NAP1	1 polypeptide; M_r 125,000	Widely distributed in eukaryotes	Responsible for binding histones H2A and H2B to DNA

Chromatin structure



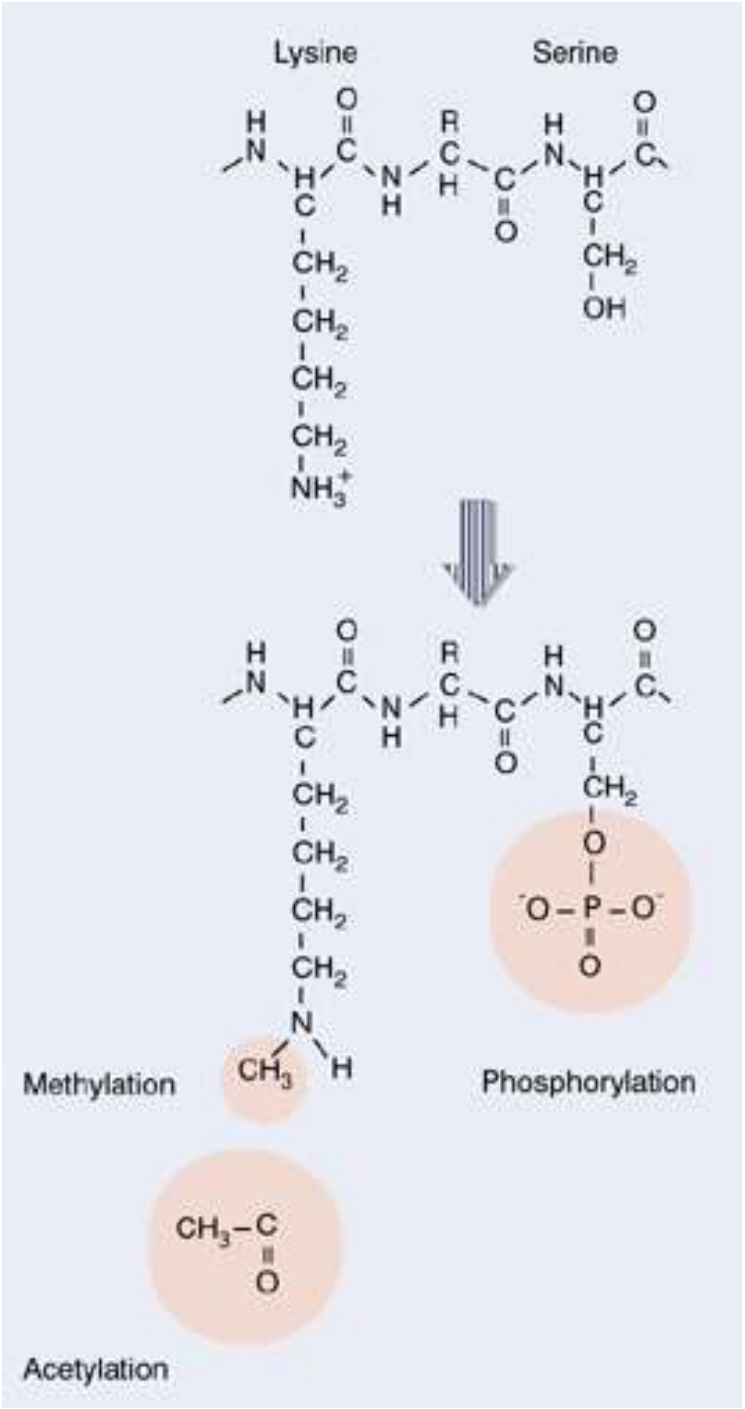
RNA Polymerase displaces DNA from the histone octamer as it advances

Not "temporarily" nucleosome free
Octamer does not dissociate to subunits
Works with covalently bound octamers

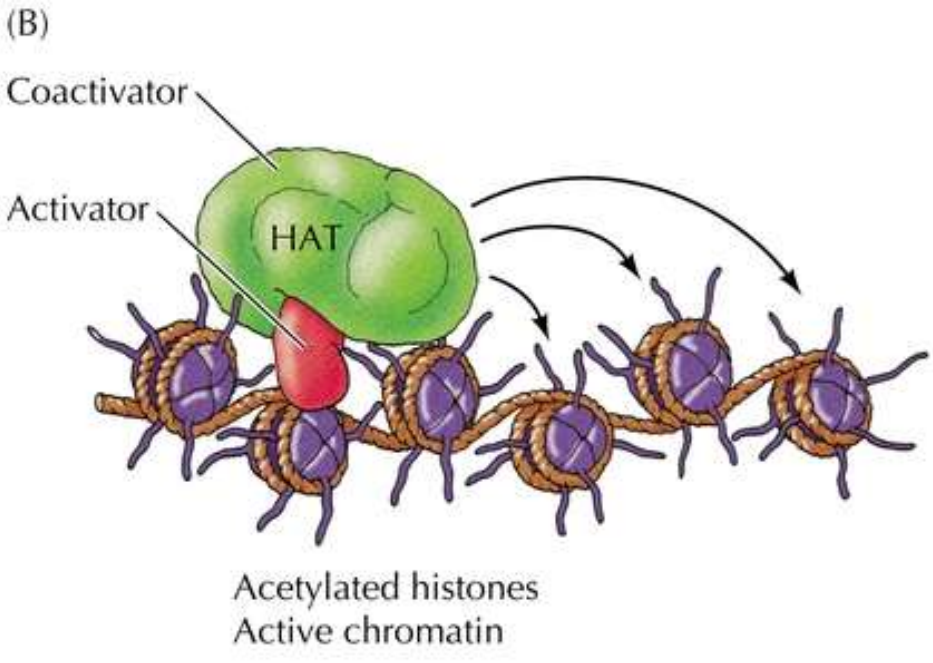
Chromatin remodeling

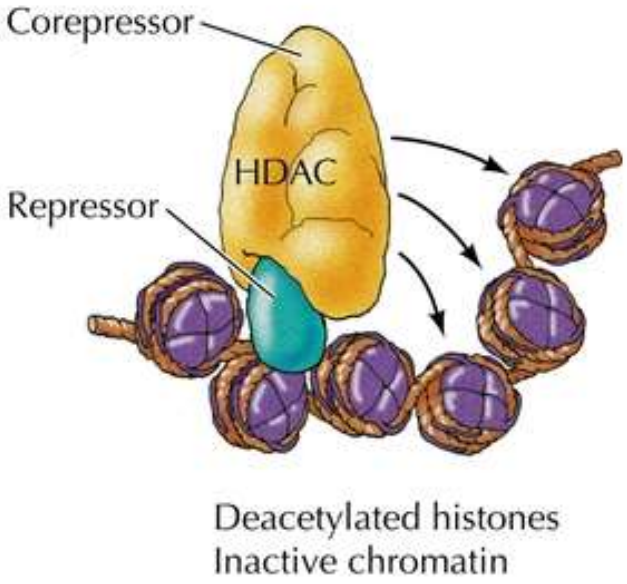
Three Major Modifications of Histone Proteins

- **acetylation**
- **phosphorylation**
- **methylation**



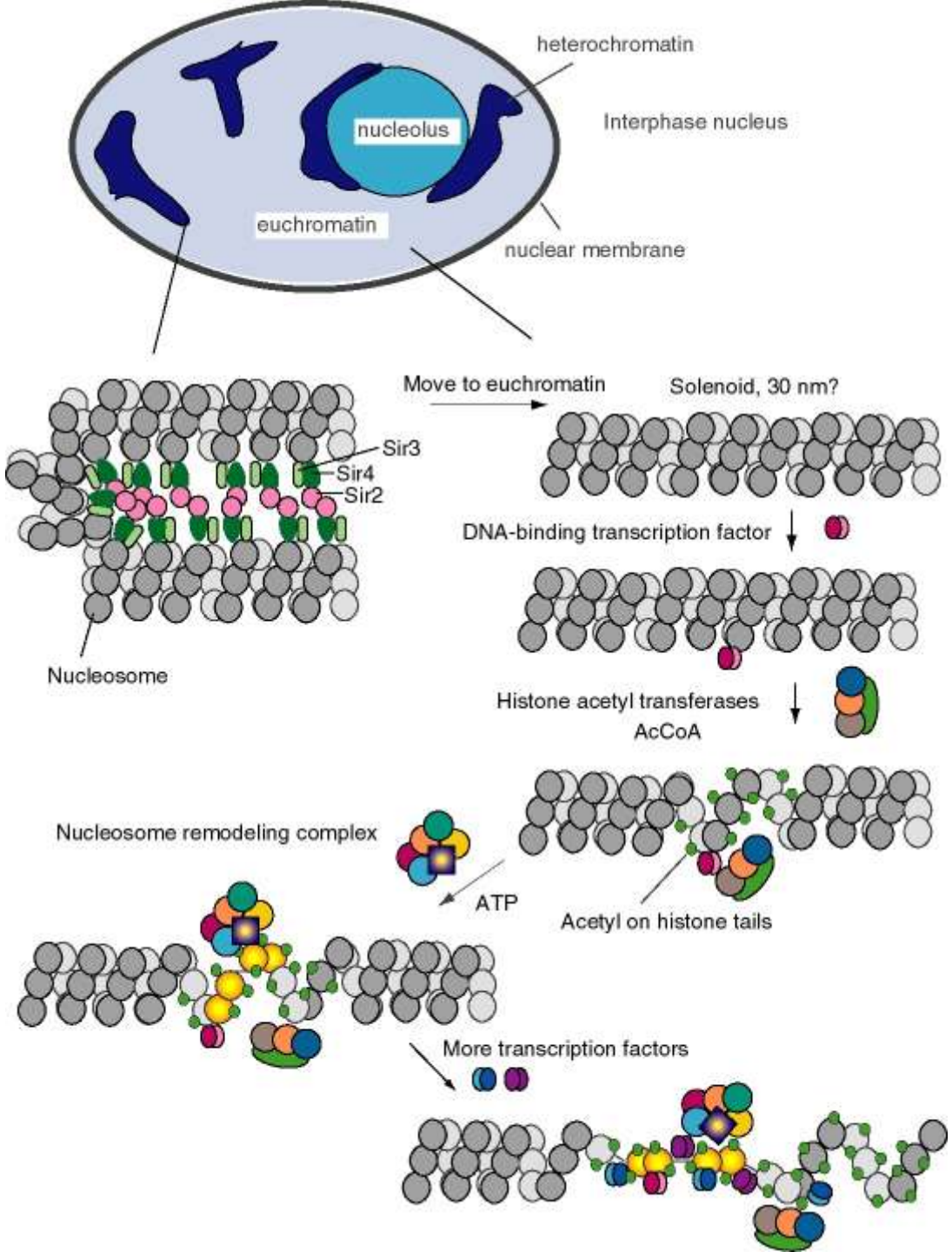
Chromatin remodeling

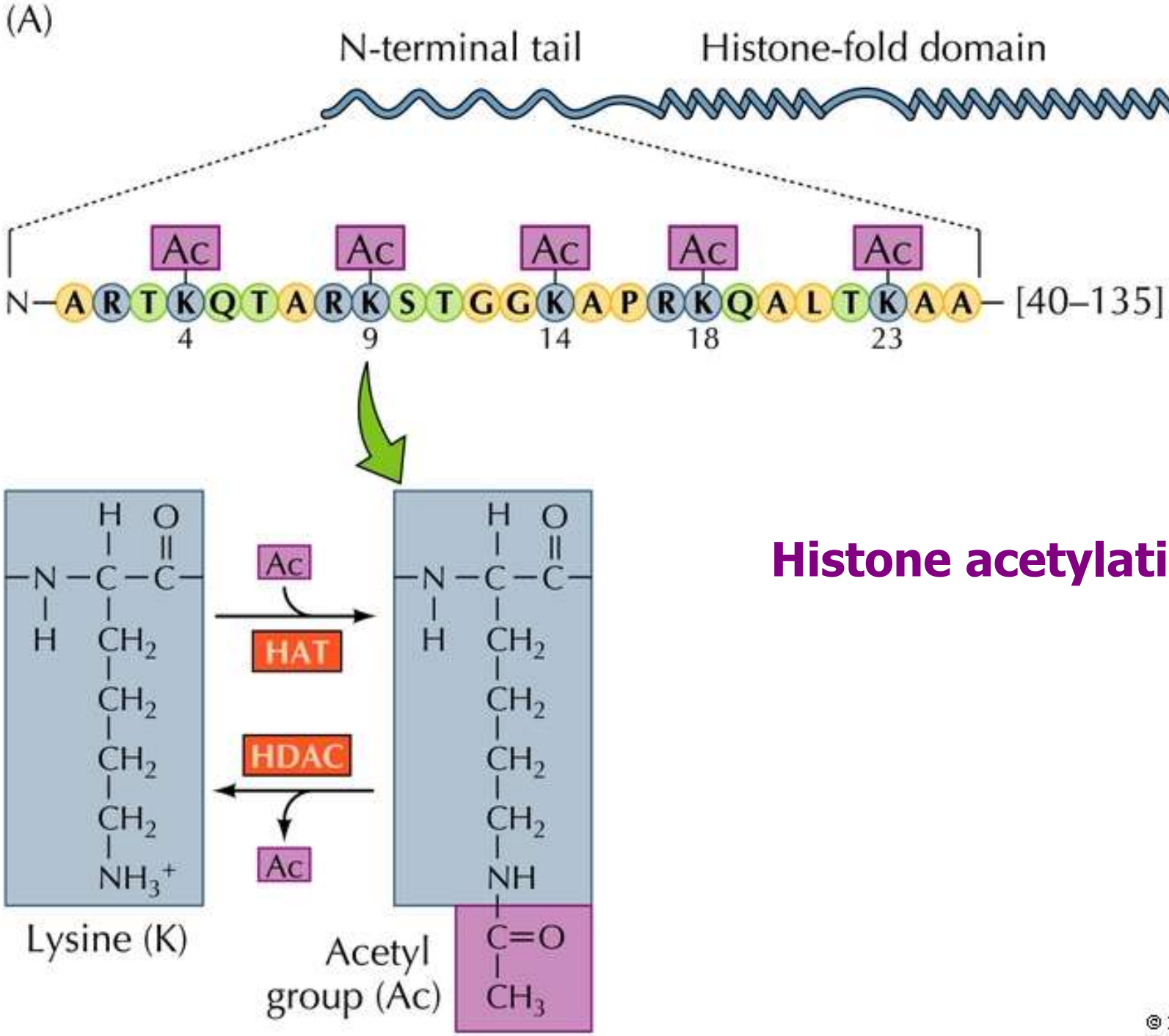




histone acetylation

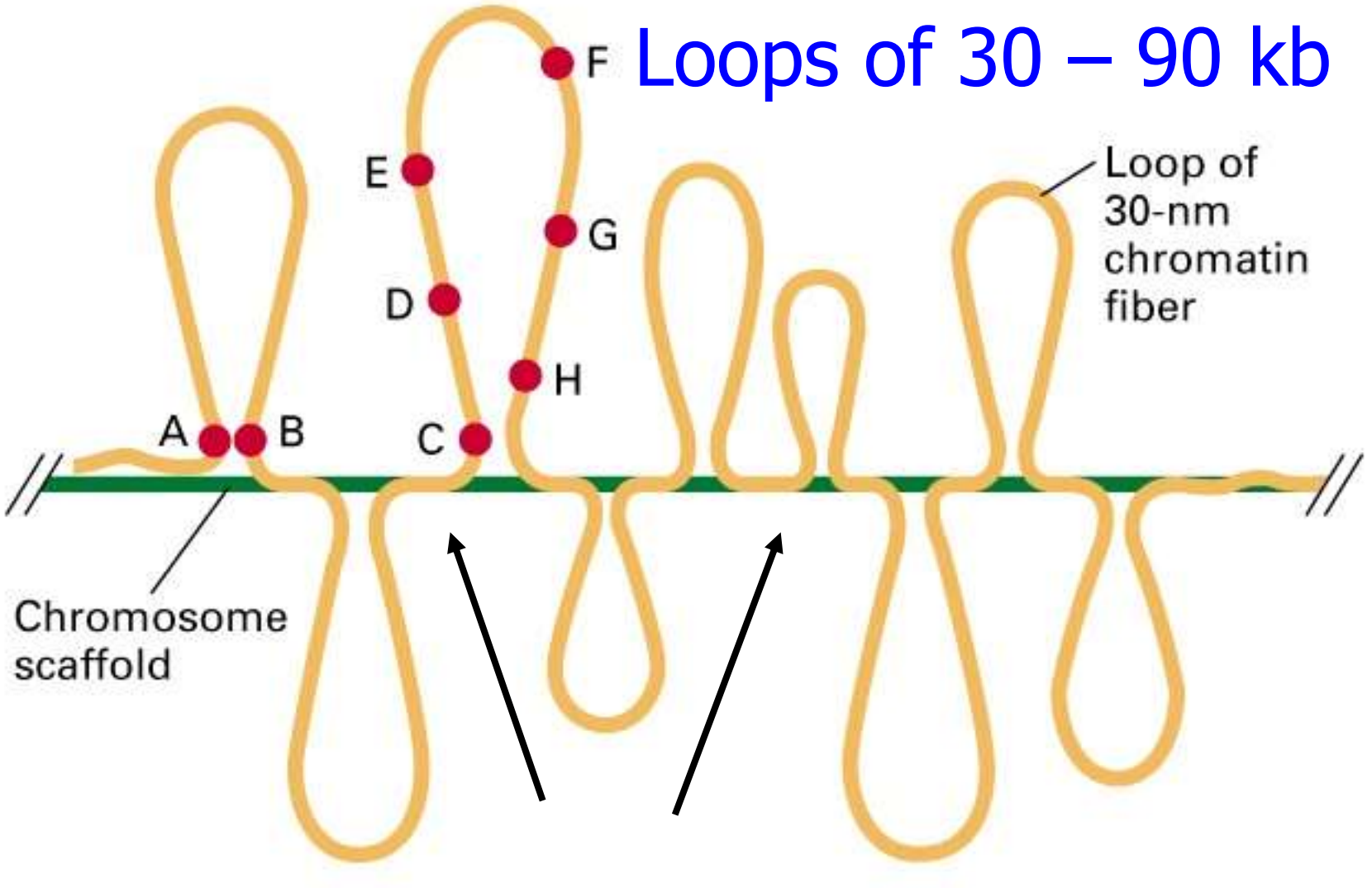
From silenced to open chromatin





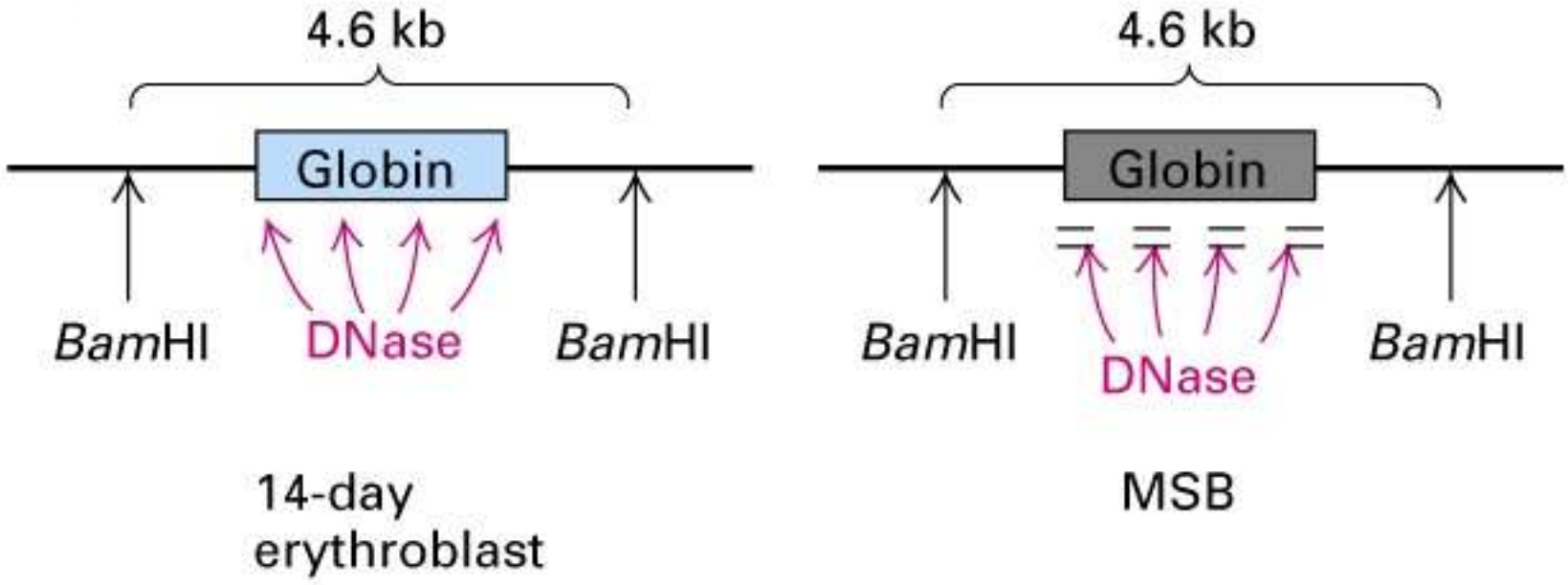
Histone acetylation

Chromatin remodeling



Chromatin structure

Actively transcribed genes are susceptible to digestion by endonucleases = *DNase I hypersensitive regions*

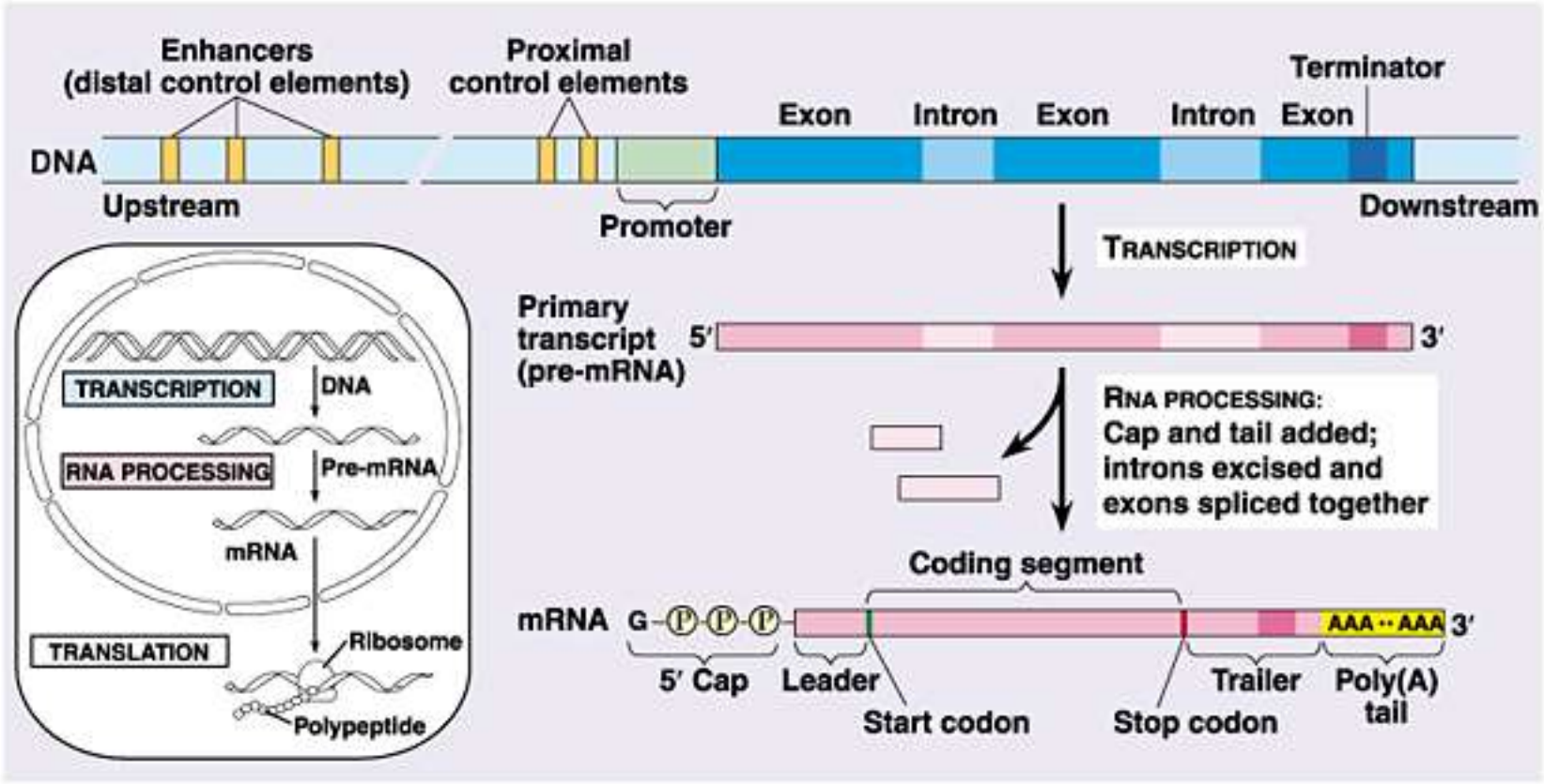


(synthesizes hemoglobin)

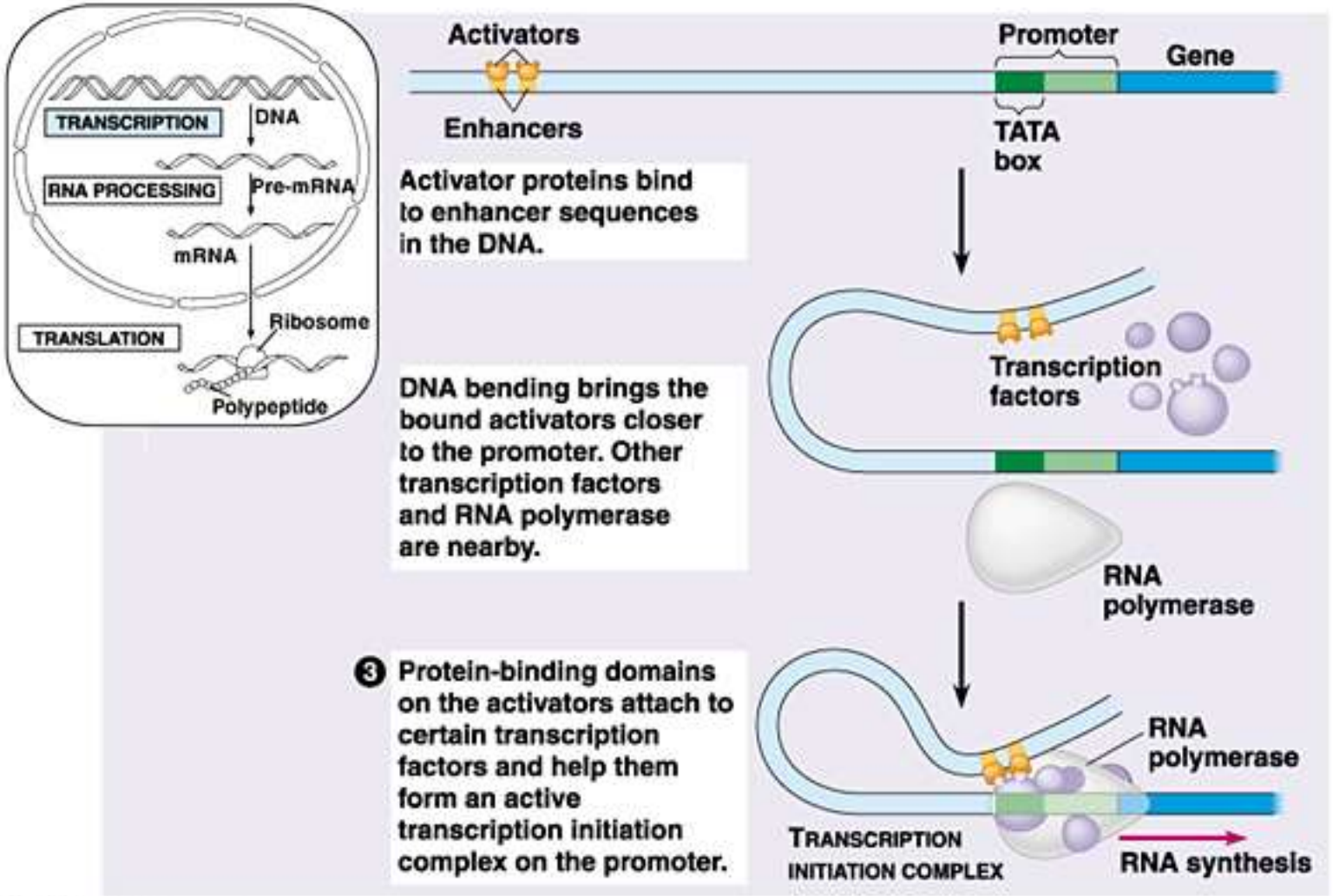
(undifferentiated cell type)

Transcription Factors

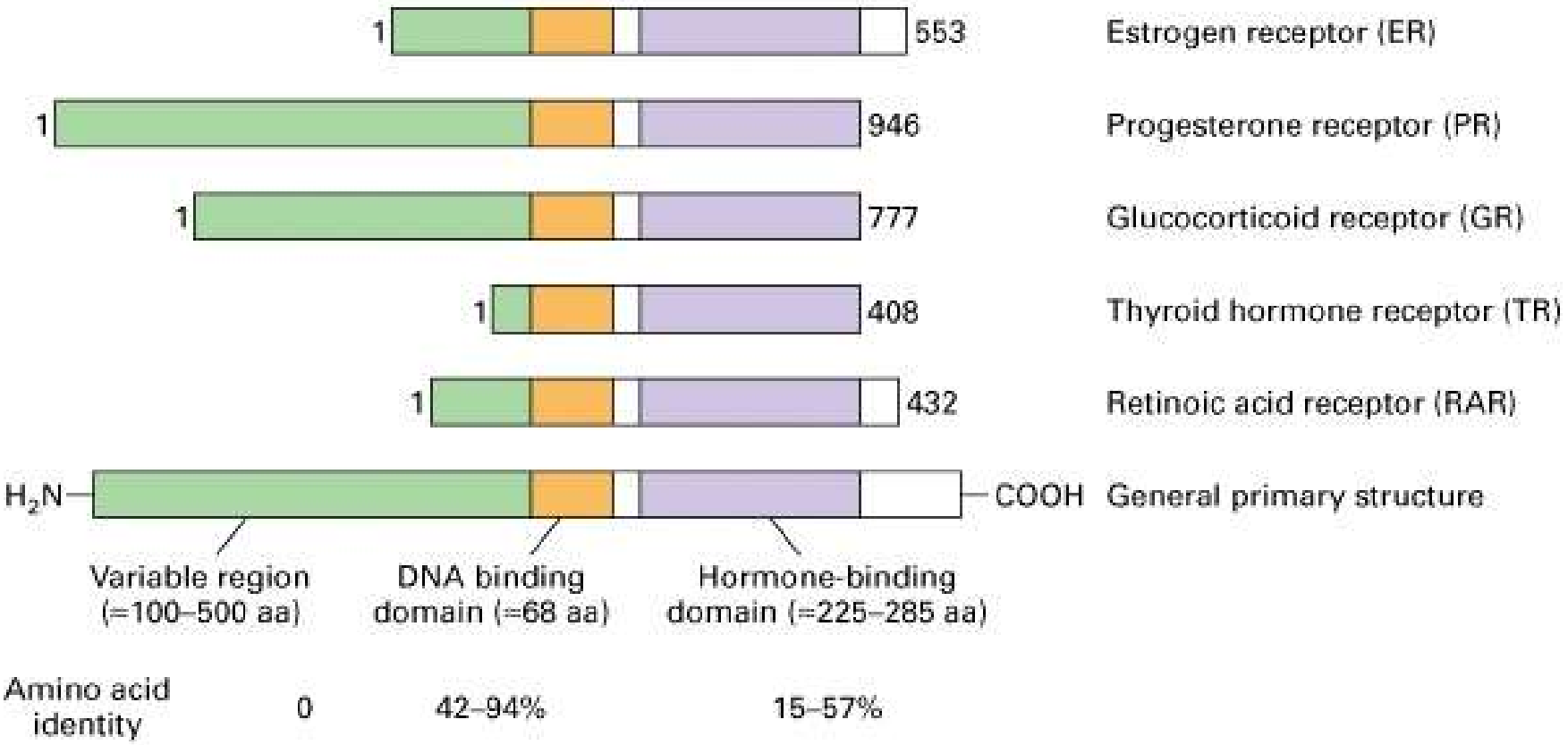
RNA polymerase cannot transcribe genes unless it is assisted by transcription factors (TFs). TFs recognize and bind specific sequences in promoters and enhancers (regions upstream of the gene being transcribed) to stimulate binding of RNA polymerase. Once a transcript is made, a 5' cap is appended, introns are spliced out and a poly(A) tail is added to the 3' end.



Transcription Factors

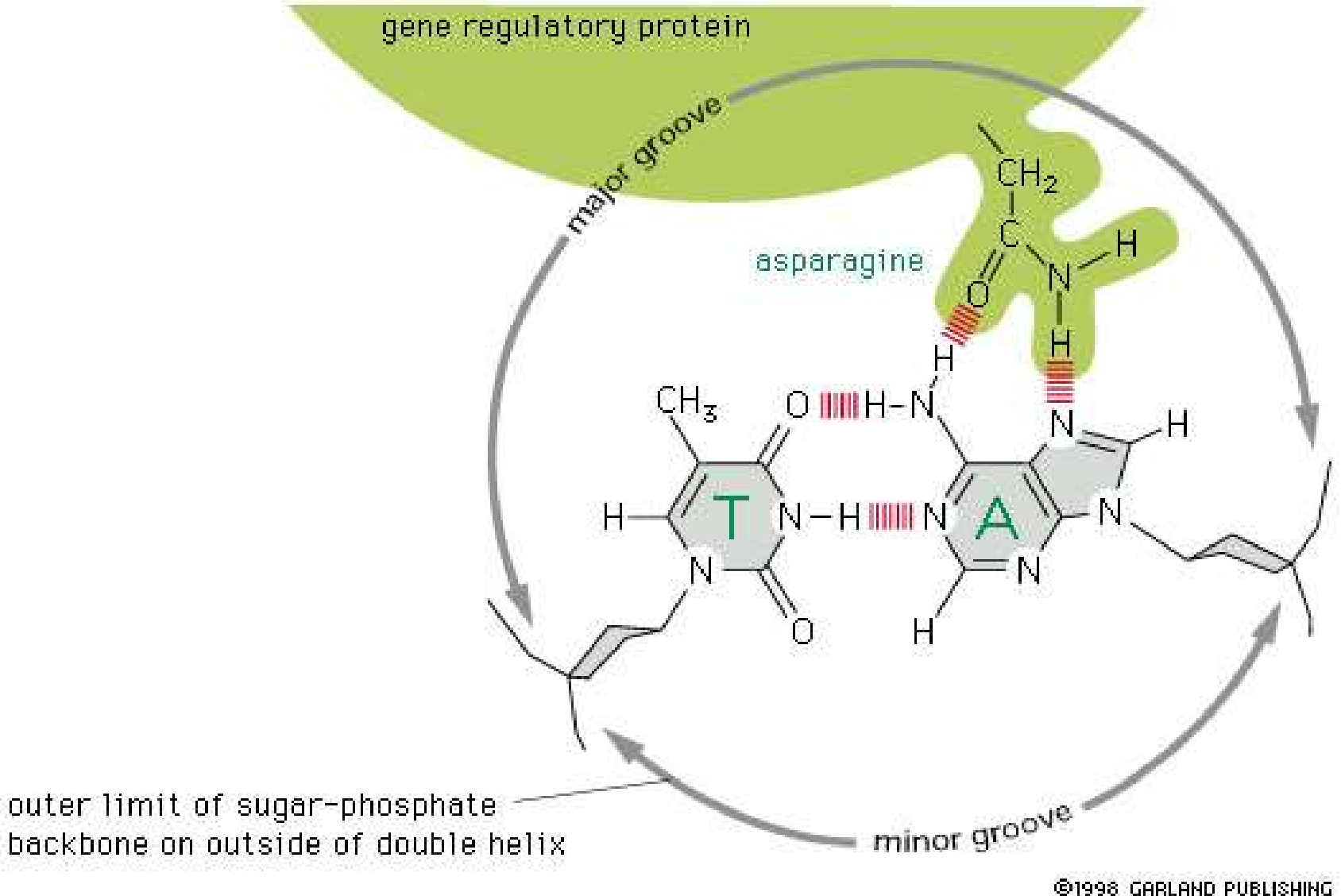


Modular nature of eukaryotic transcription factors

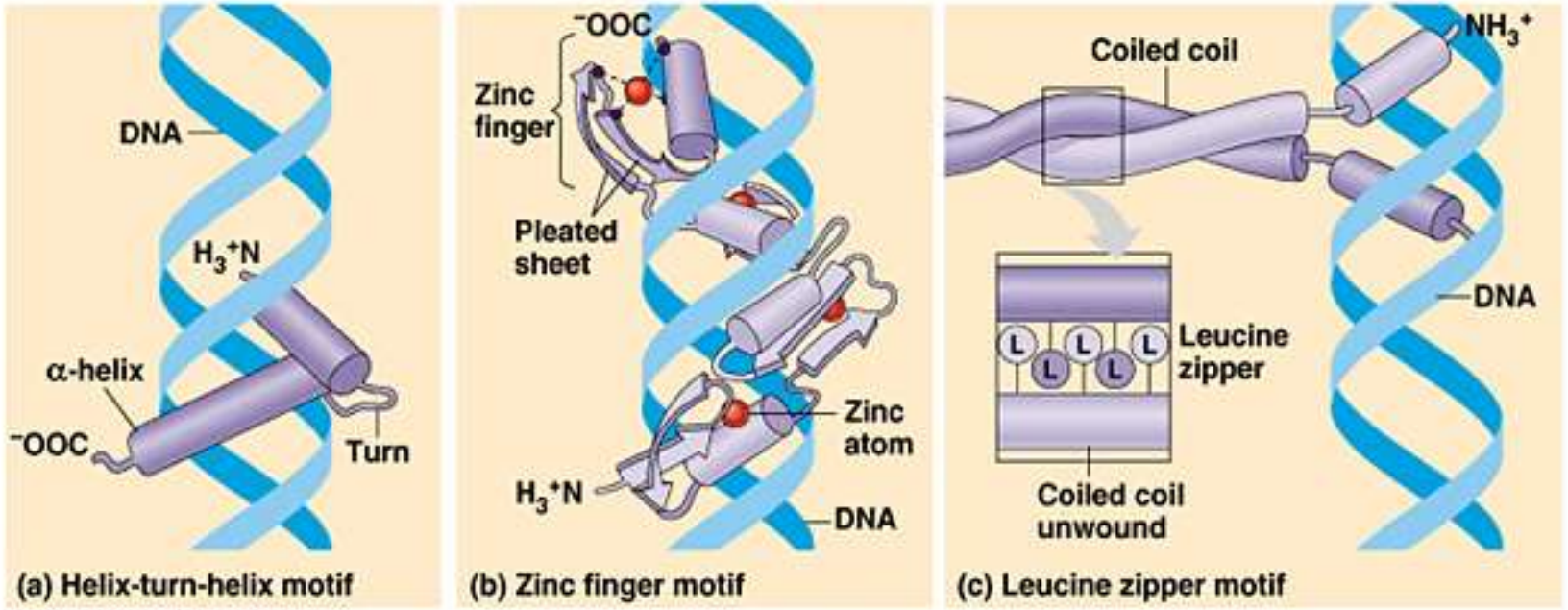


Every transcription factor protein contains at least one, often more, activation domain and one DNA-binding domain. Many transcription factors bind to DNA as dimers recognizing palindromic sequences.

Transcription factors bind to DNA

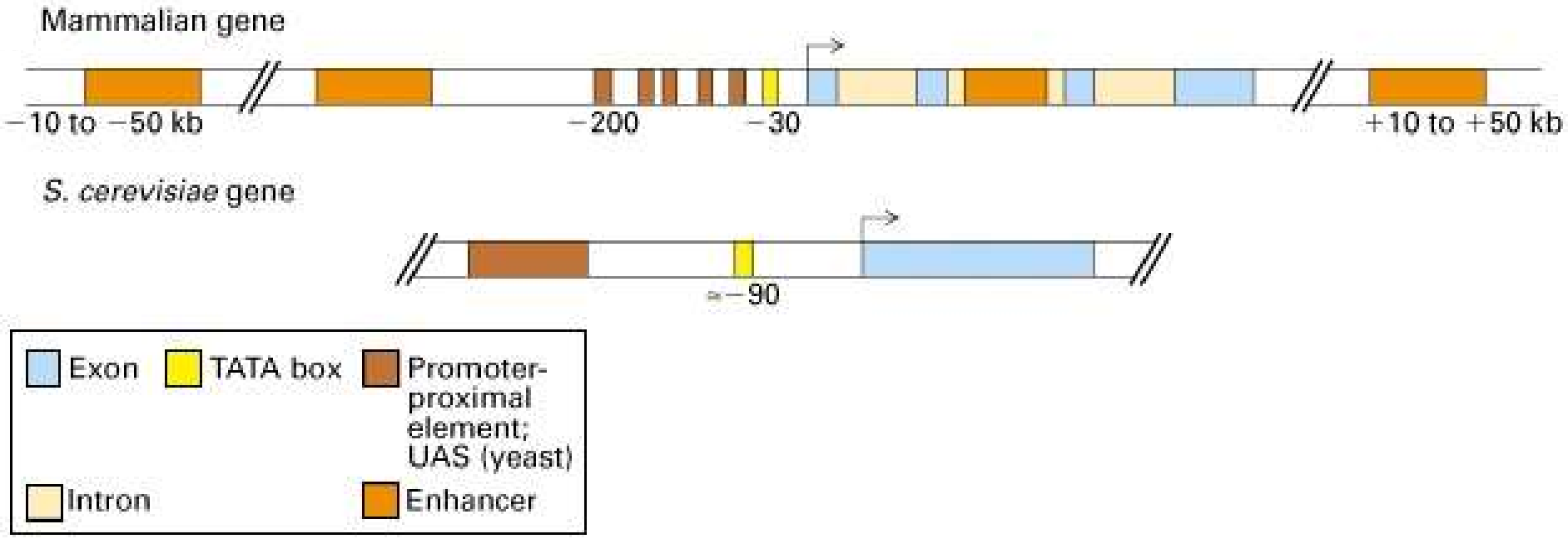


DNA binding domains

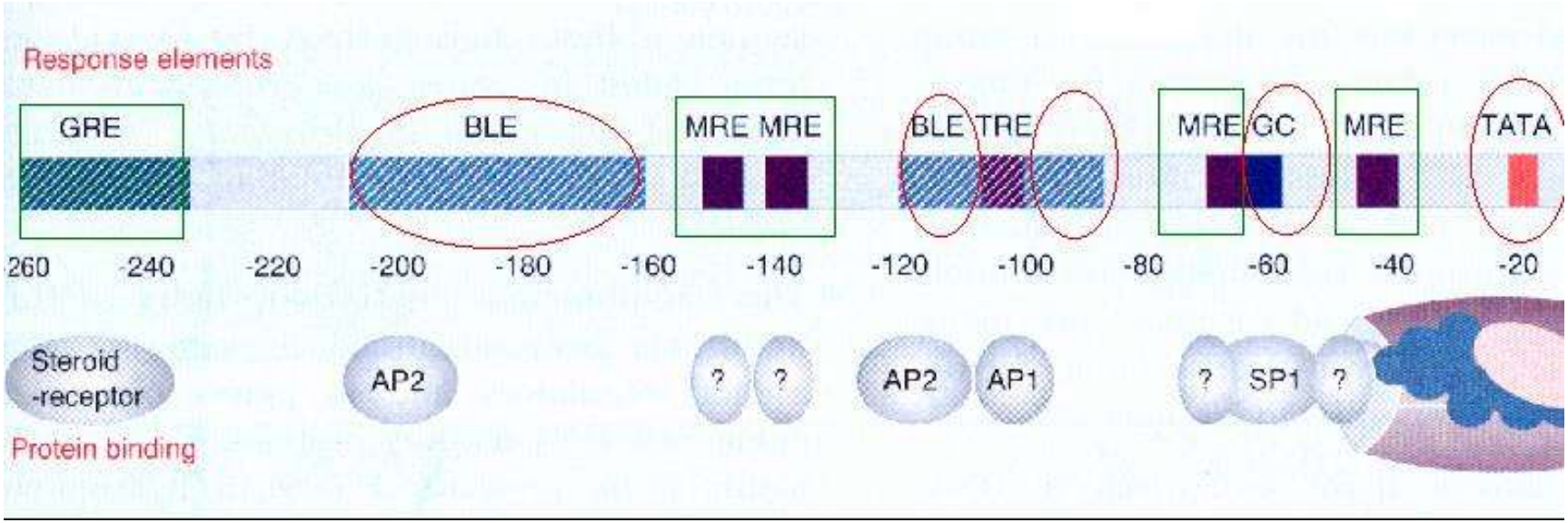


There are many strategies that proteins use to bind DNA in a highly specific manner. HTH, zinc finger, and leucine zippers are three common motifs. Each motif relies on making numerous hydrogen bonds, ionic bonds and van der wa'als contacts with the bases and sugar-phosphate backbone.

Molecular anatomy of eukaryotic genes



Regulatory elements of Human Metallothionein Gene



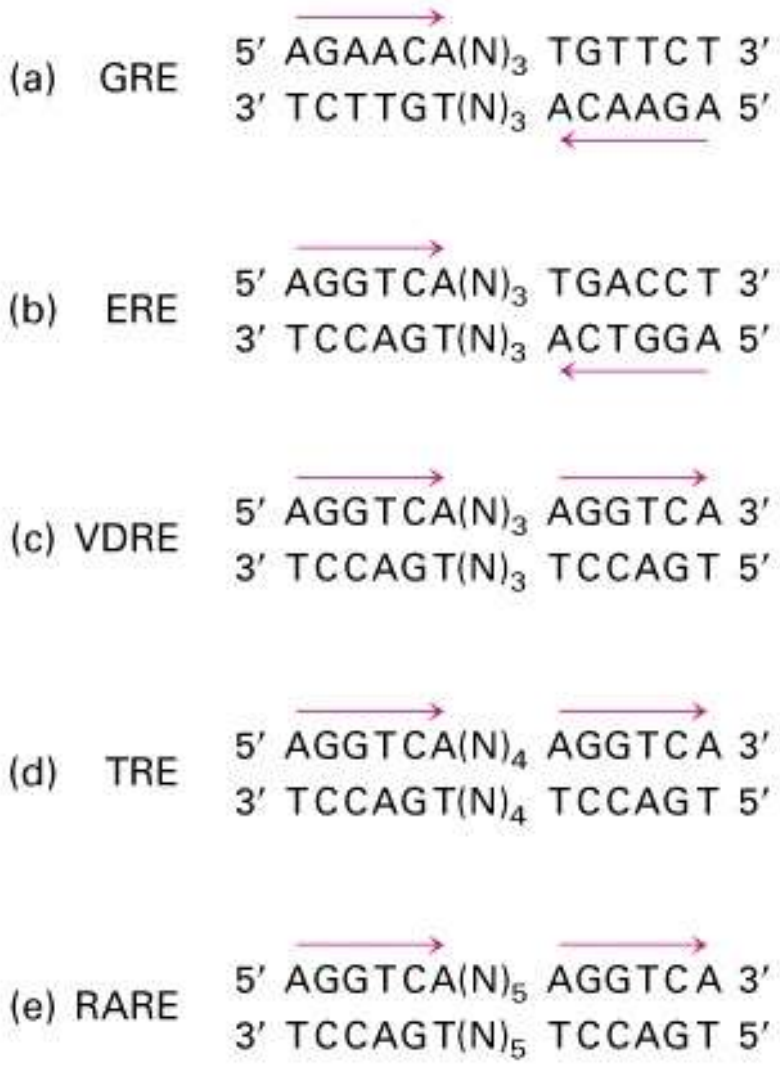
UASs - Upstream Factors – constituvely exprimed

Response Elements – induced (regulated)

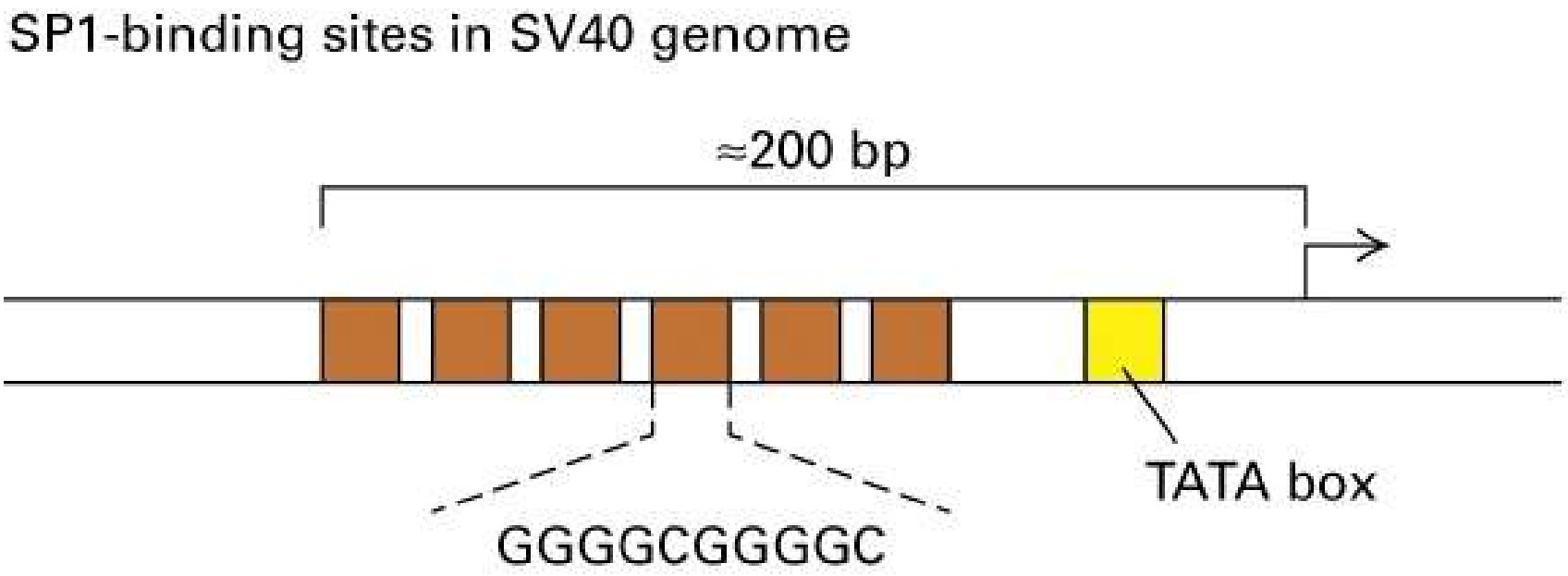
Response elements for transcription factors

- ***Inverted repeat*** **structures of enhancers for transcription factors reflect homodimeric subunit organization**

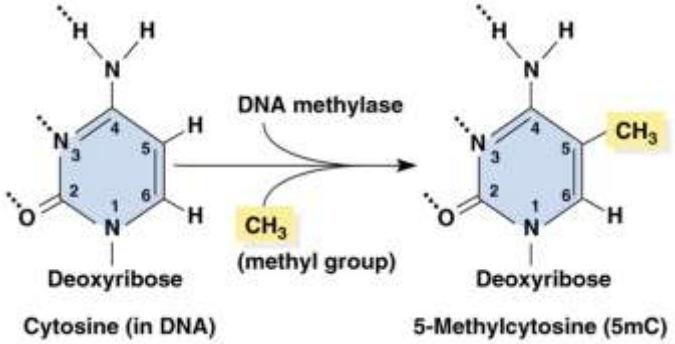
- ***Direct repeat*** **structures reflect heterodimeric subunit organization**

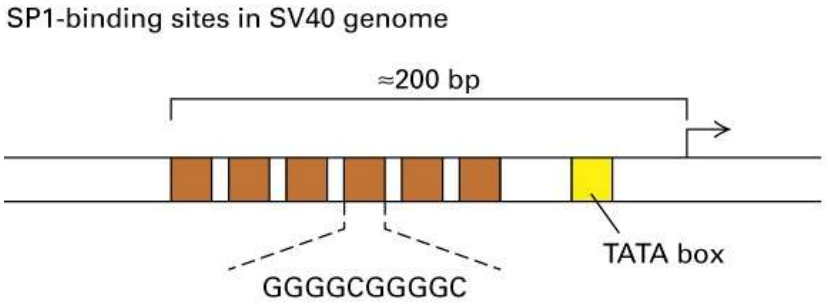


SP1 is a transcription factor that binds to CpG UAS

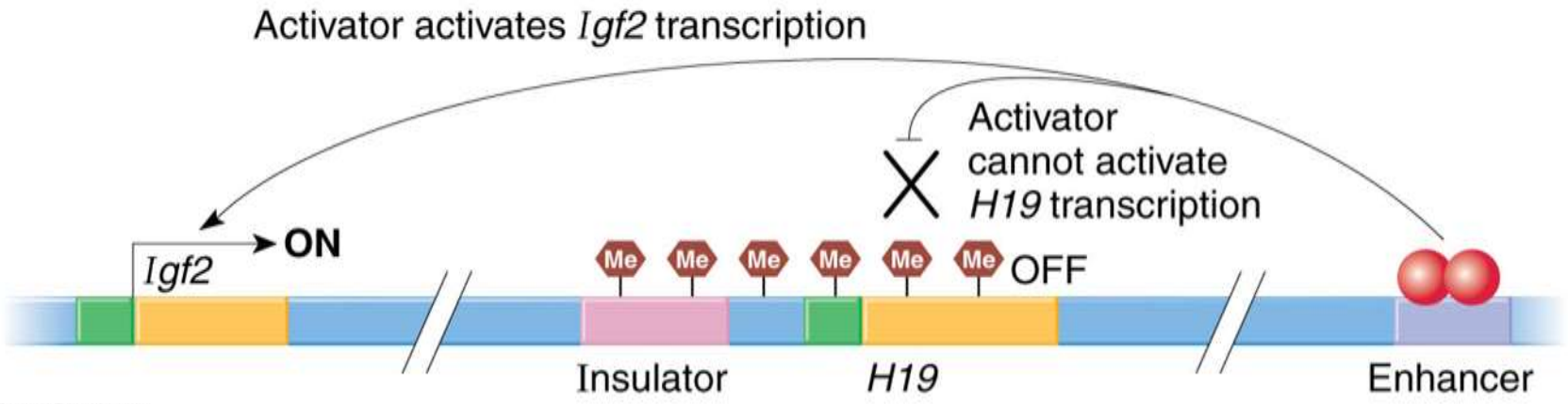


SP1 is a transcription factor that binds to CpG UAS



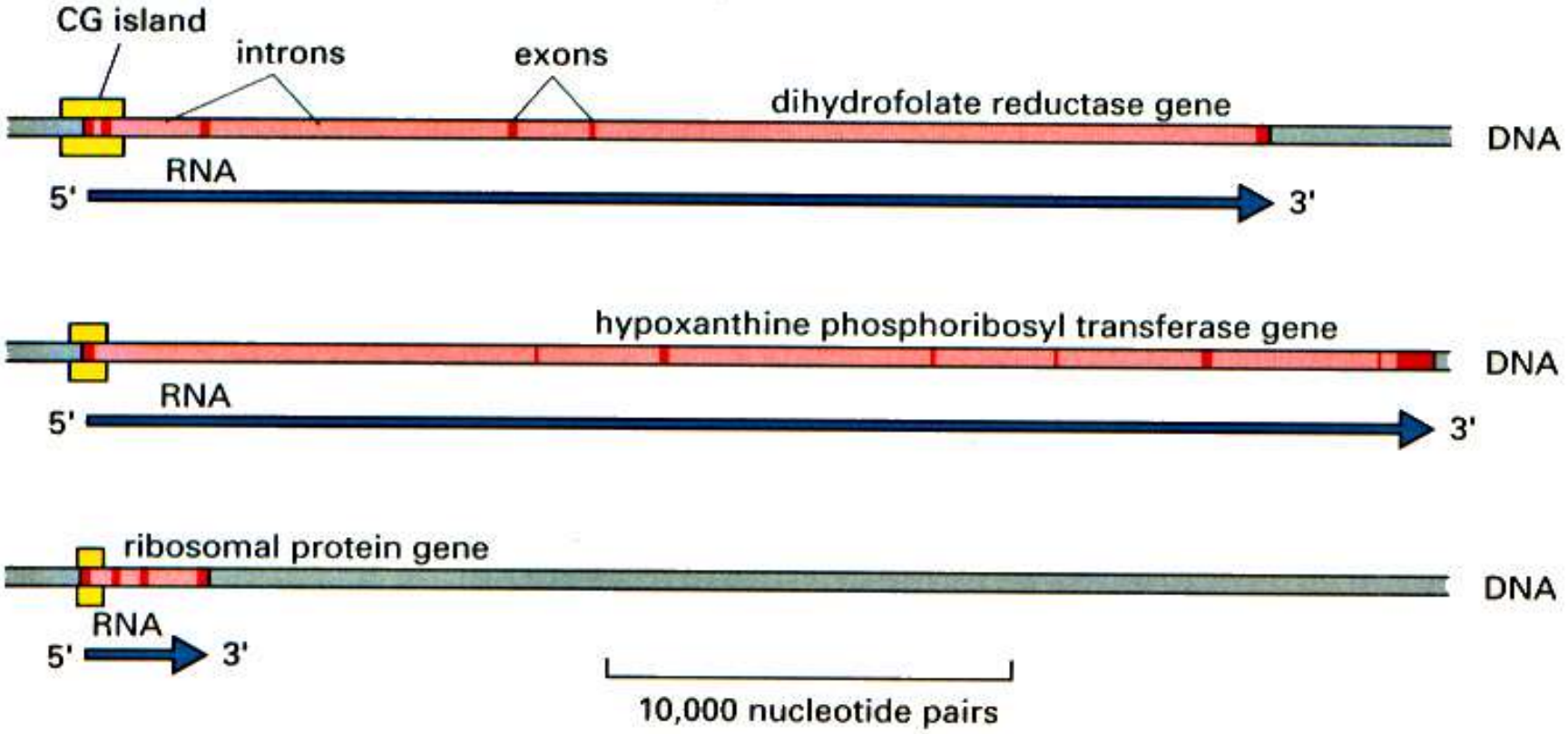


b) Paternal chromosome

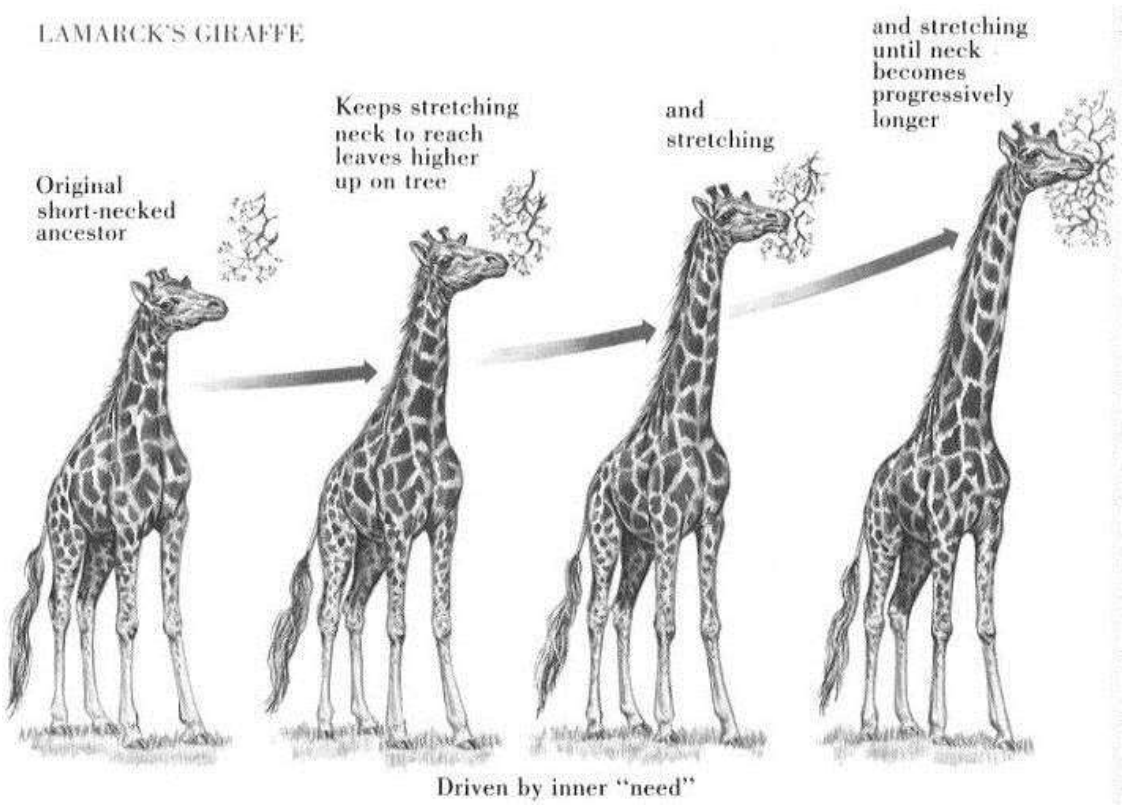


CG islands surrounding the promoters of three mammalian genes

CG sequences are unevenly distributed in the genome. They are clustered in islands, about 1000 to 2000 nucleotides long, near promoters of genes. The CG islands of active genes are kept in an demethylated state. In inactive genes, the Cs are methylated



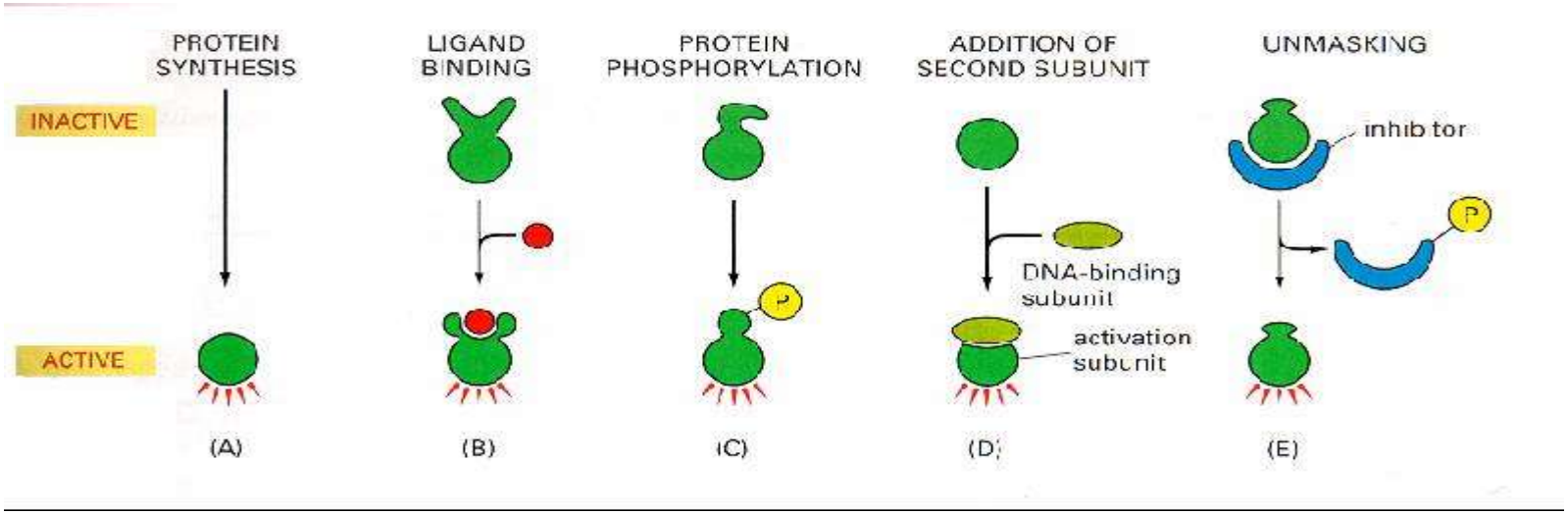
Chromatin remodeling and DNA methylation are the basis for epigenetic inheritance.



Response elements

Response elements are recognised and bind by induction factors

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCCNNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF



Enhancers

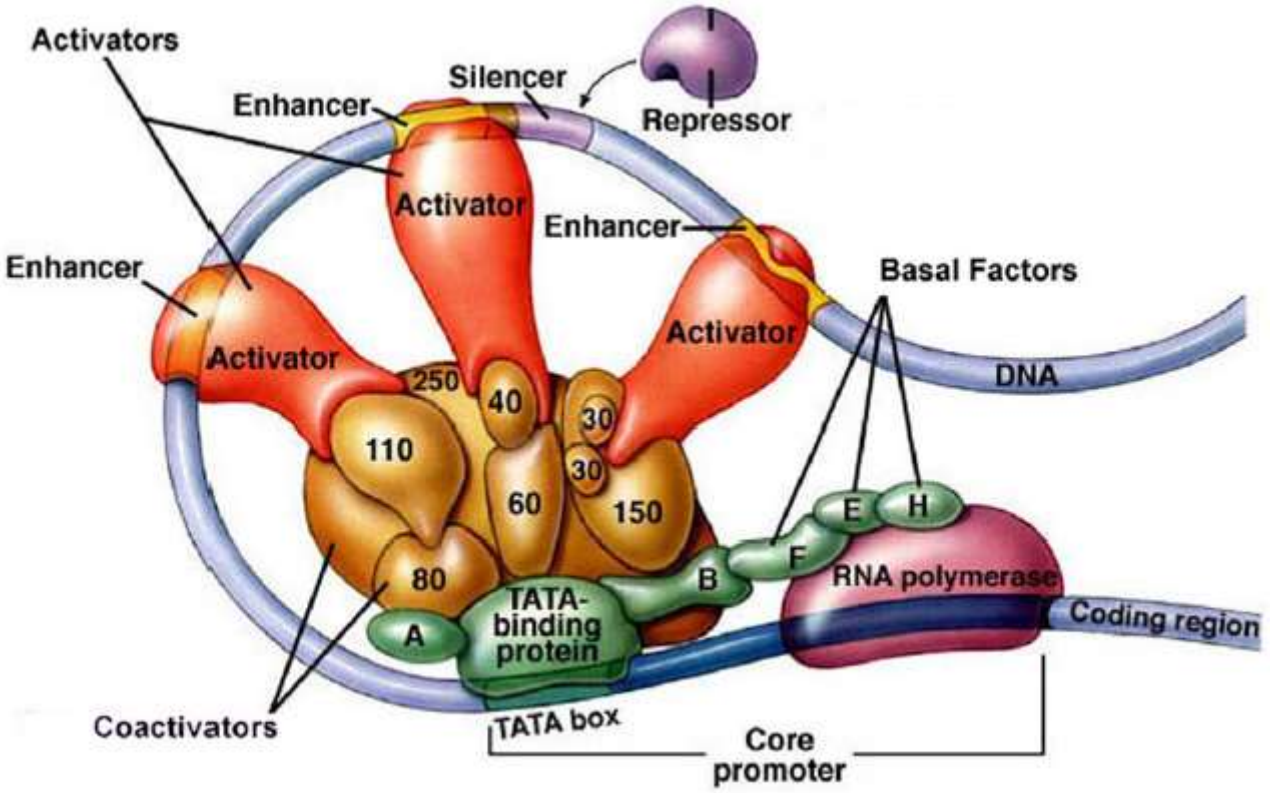
- **Enhancers act at a distance from the transcription unit that they act on**
- **Enhancers are usually found upstream, but they can be present in introns or downstream of a gene**
- **Enhancers act in either orientation without affecting the direction of transcription**
- **Enhancers are usually complex, e.g. they are composed of multiple protein binding sites**
- **Average length of an enhancer 50-200 bp**
- **The enhancer sequence is irrelevant to enhancer action as long as the transcription factor(s) can bind with appropriate affinity**

How do enhancers act from a distance?

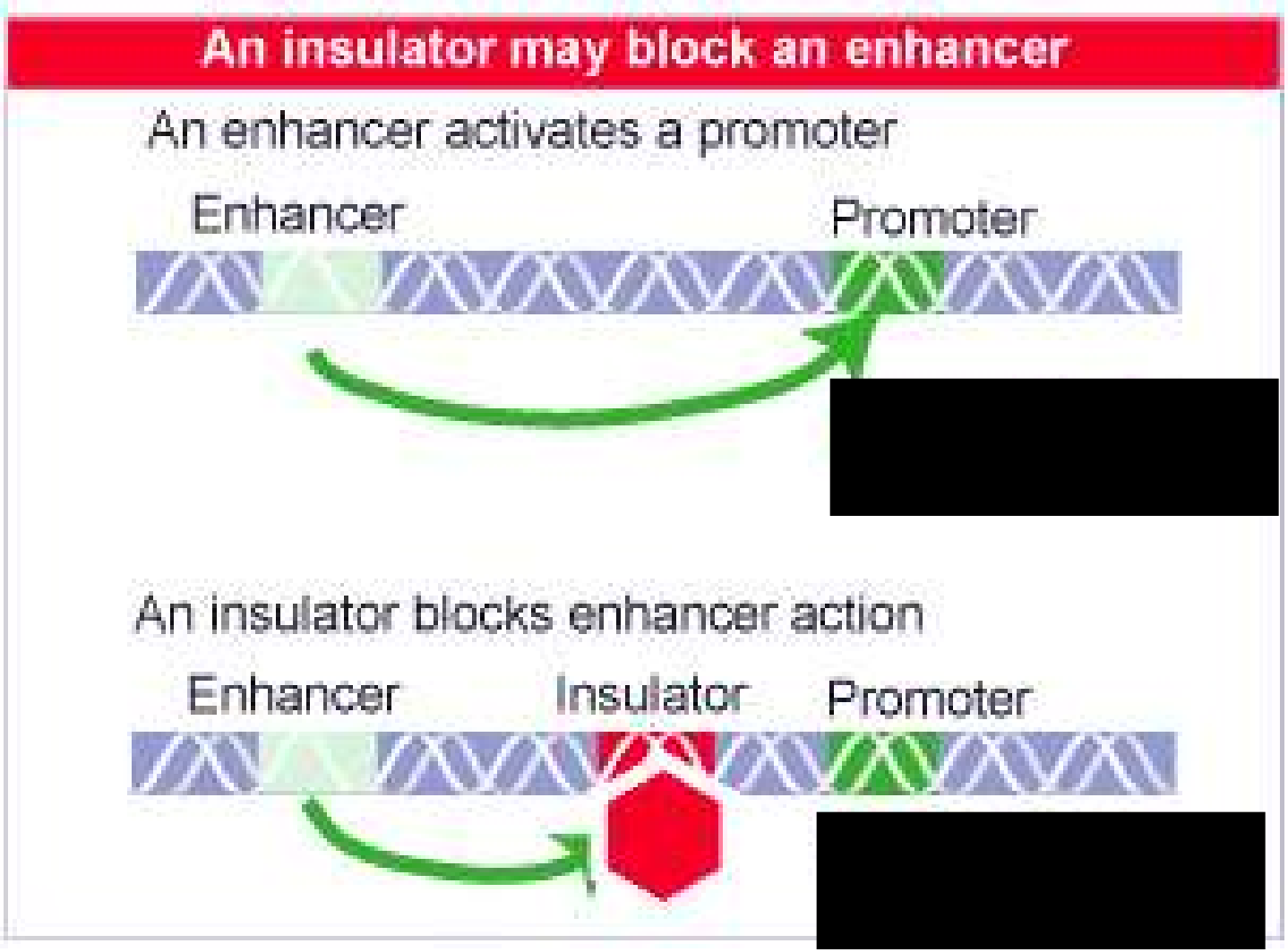
1. **The "conformational model": binding of a transcription factor to its' enhancer causes chromatin structure to change, resulting in better accessibility of the promoter to RNApolII**
2. **The "scanning model": binding of a transcription factor to its' enhancer causes it to assume an "*activated state*", which is conveyed to RNApolII when it is encountered near the promoter**
3. **The "loop model": DNA looping brings the transcription factor bound to its' enhancer closer to RNApolII sitting at the promoter**

The "loop model" is favored by most researchers

How do enhancers act from a distance?



Insulators

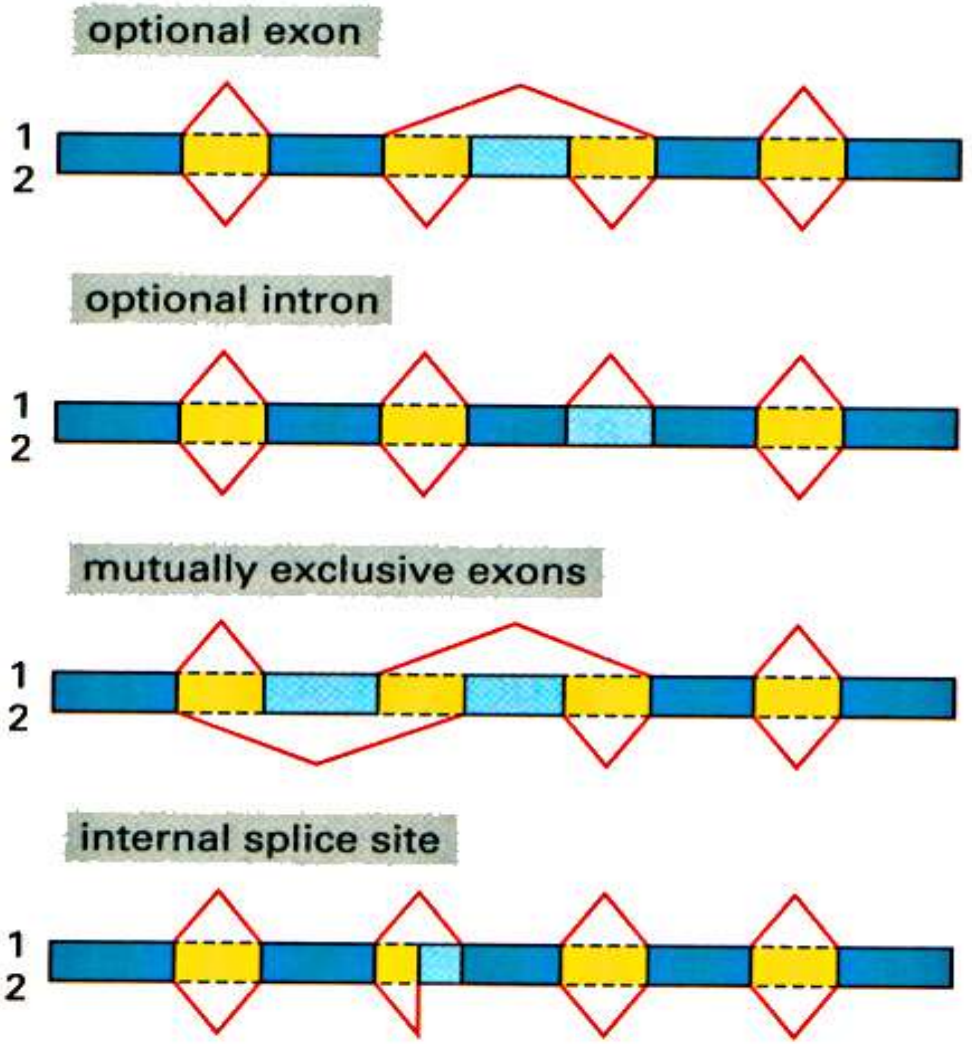

An insulator may block an enhancer
An enhancer activates a promoter
Enhancer
Promoter
An insulator blocks enhancer action
Enhancer
Insulator
Promoter

Alternative splicing is another method to regulate gene expression

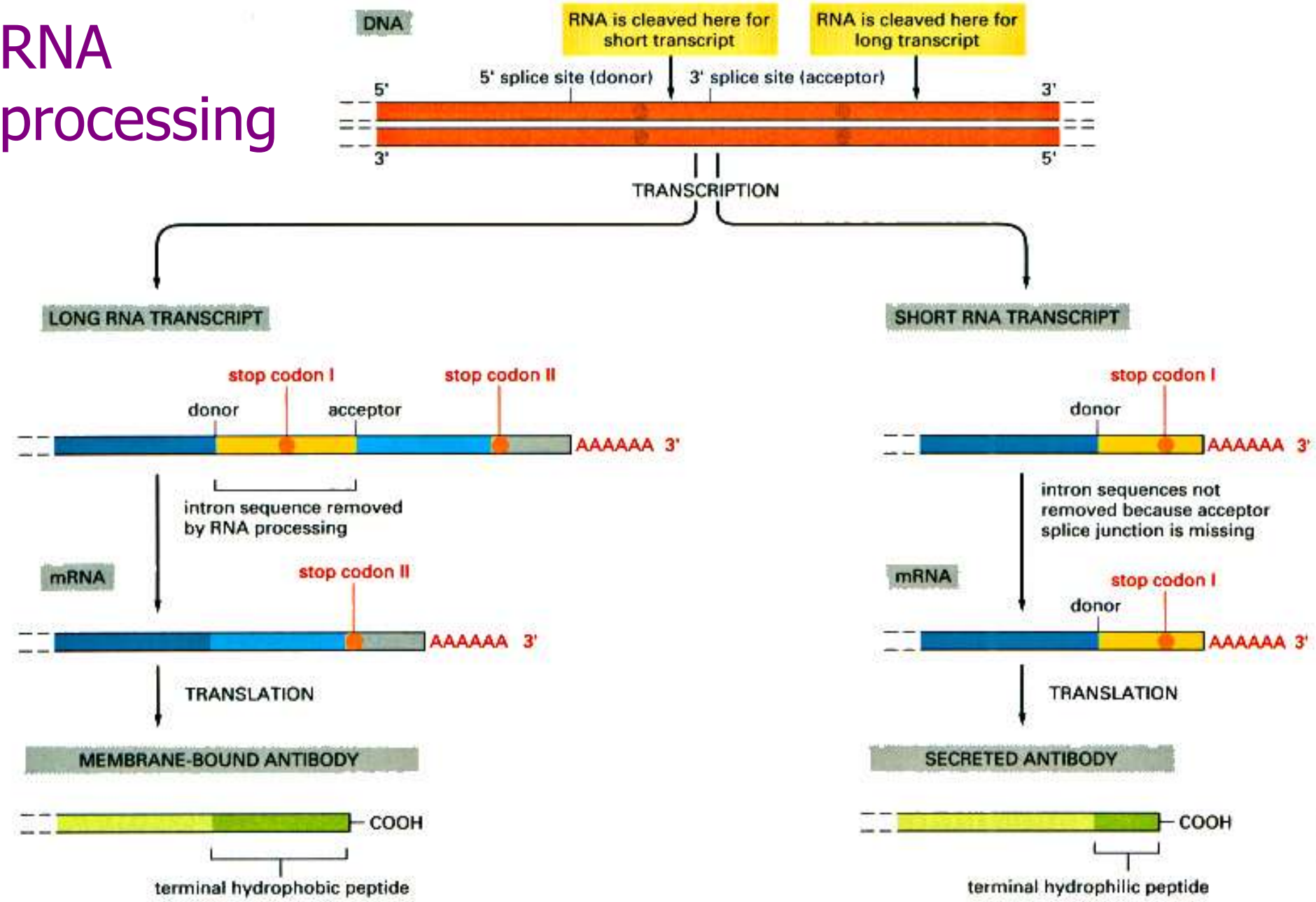
Dark blue boxes mark exons that are retained in both mRNAs.

Light blue boxes mark possible exon sequences that are included in only one of the mRNAs.

The boxes are joined by red lines to indicate where intron sequences (yellow) are removed.

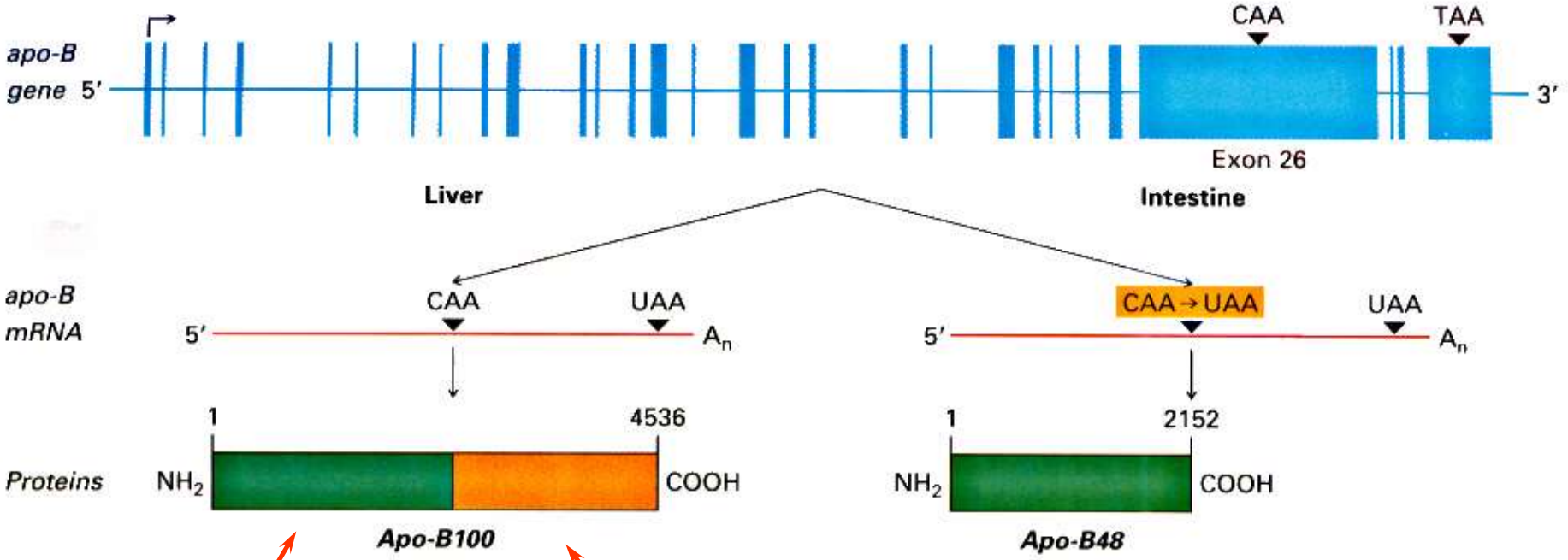


RNA processing

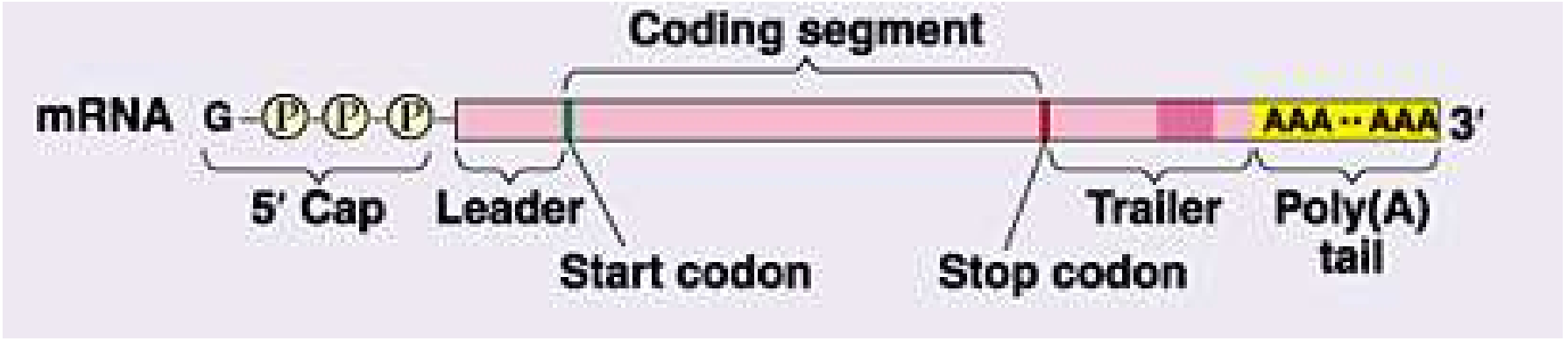


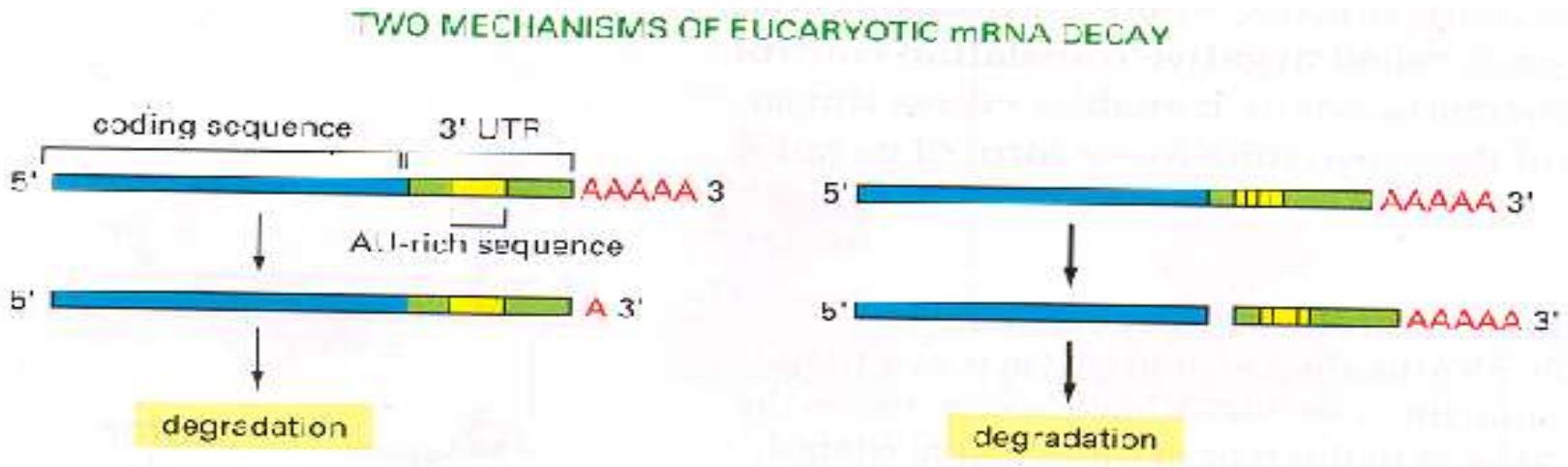
RNA editing of human lipoprotein apo-B

Apo-B mRNA produced in the human liver is translated into the apo-B100 protein In intestinal cells, however, the same mRNA is edited, deaminating a C to a U. This converts a Gln codon into a stop codon. Translation of this edited mRNA produces apo-B48, a protein that lacks the C-terminal portion of apo-B100.



mRNA Degradation

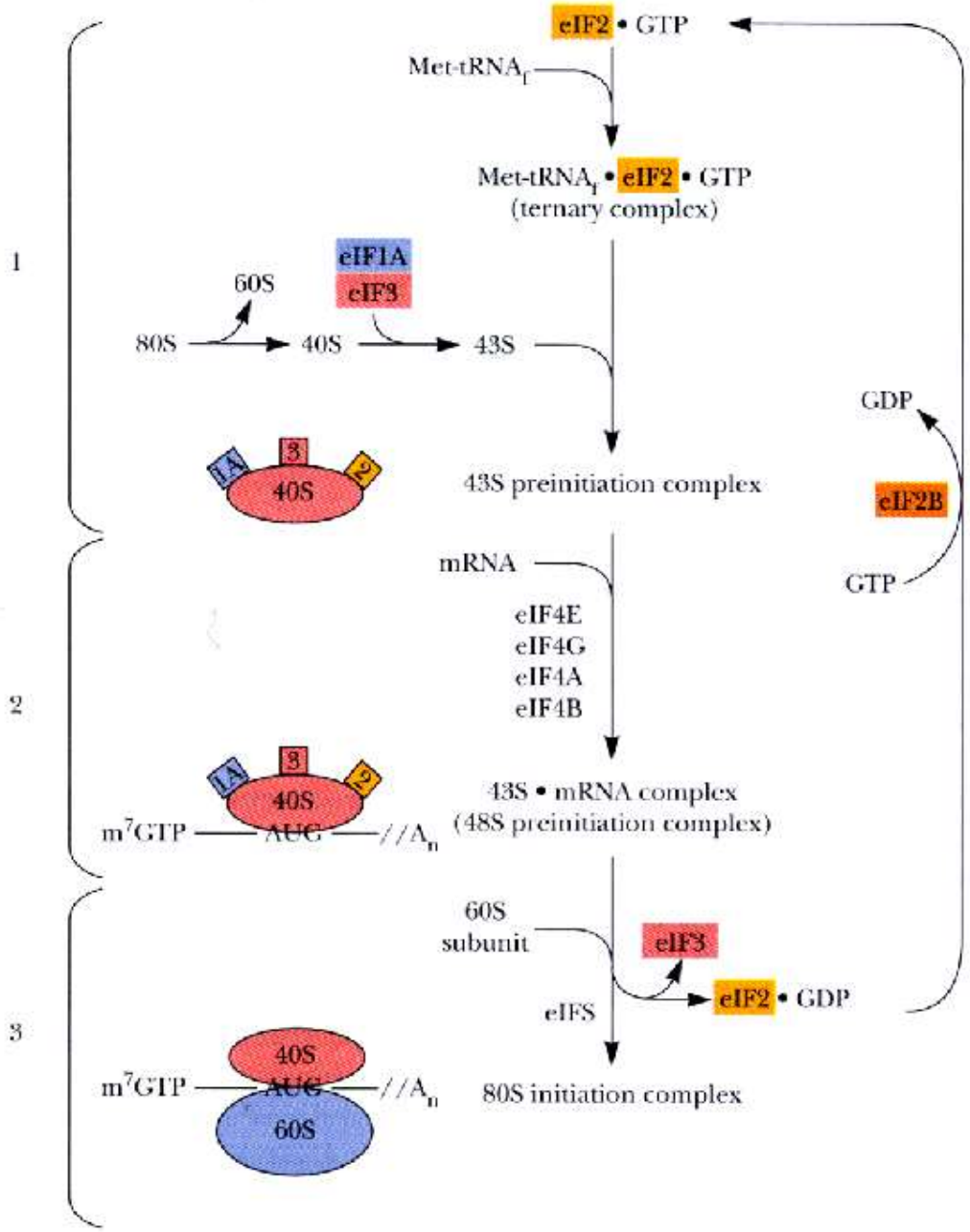




mRNA stablility is greatly reduced when the 5’ cap and 3’ poly(A) modifications are removed. There is a correlation between the length of the poly(A) tail and mRNA stability.

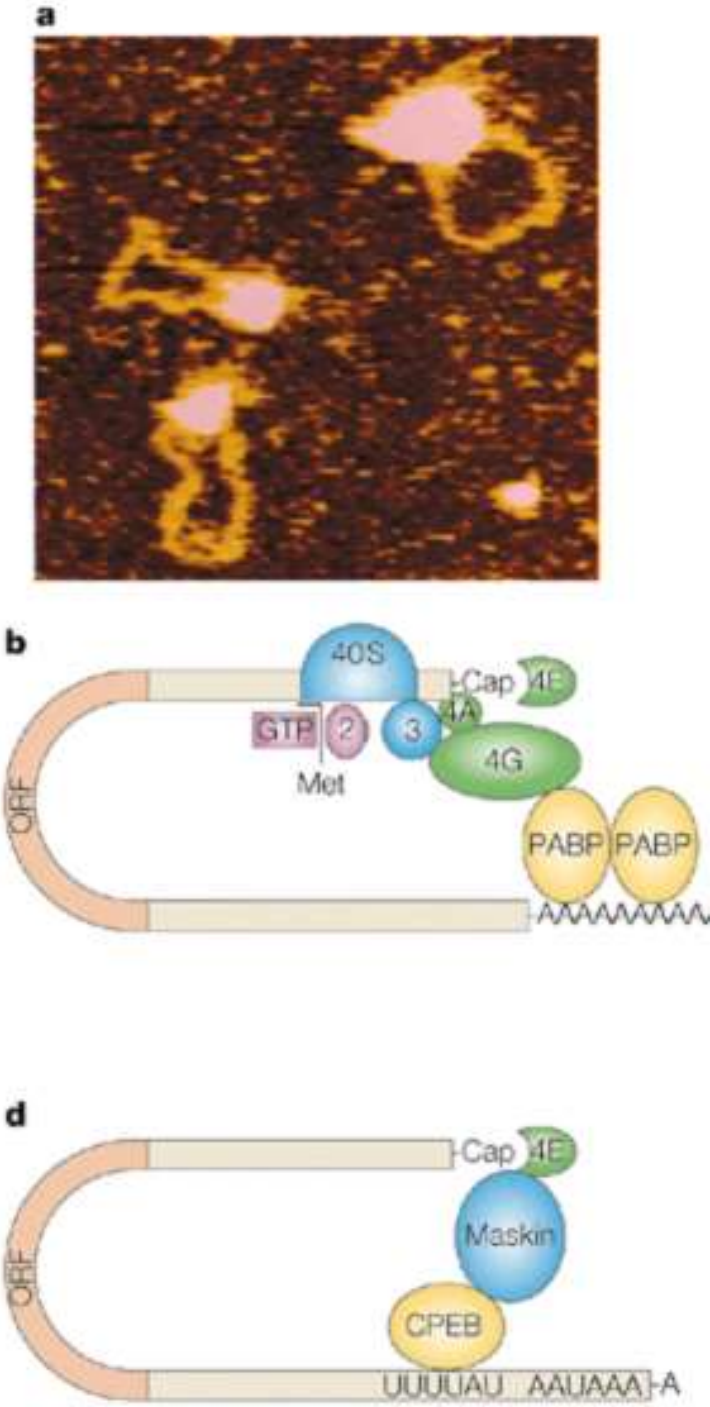
Regulation at Transcription level

eIF2 is blocked in GDP bounded state



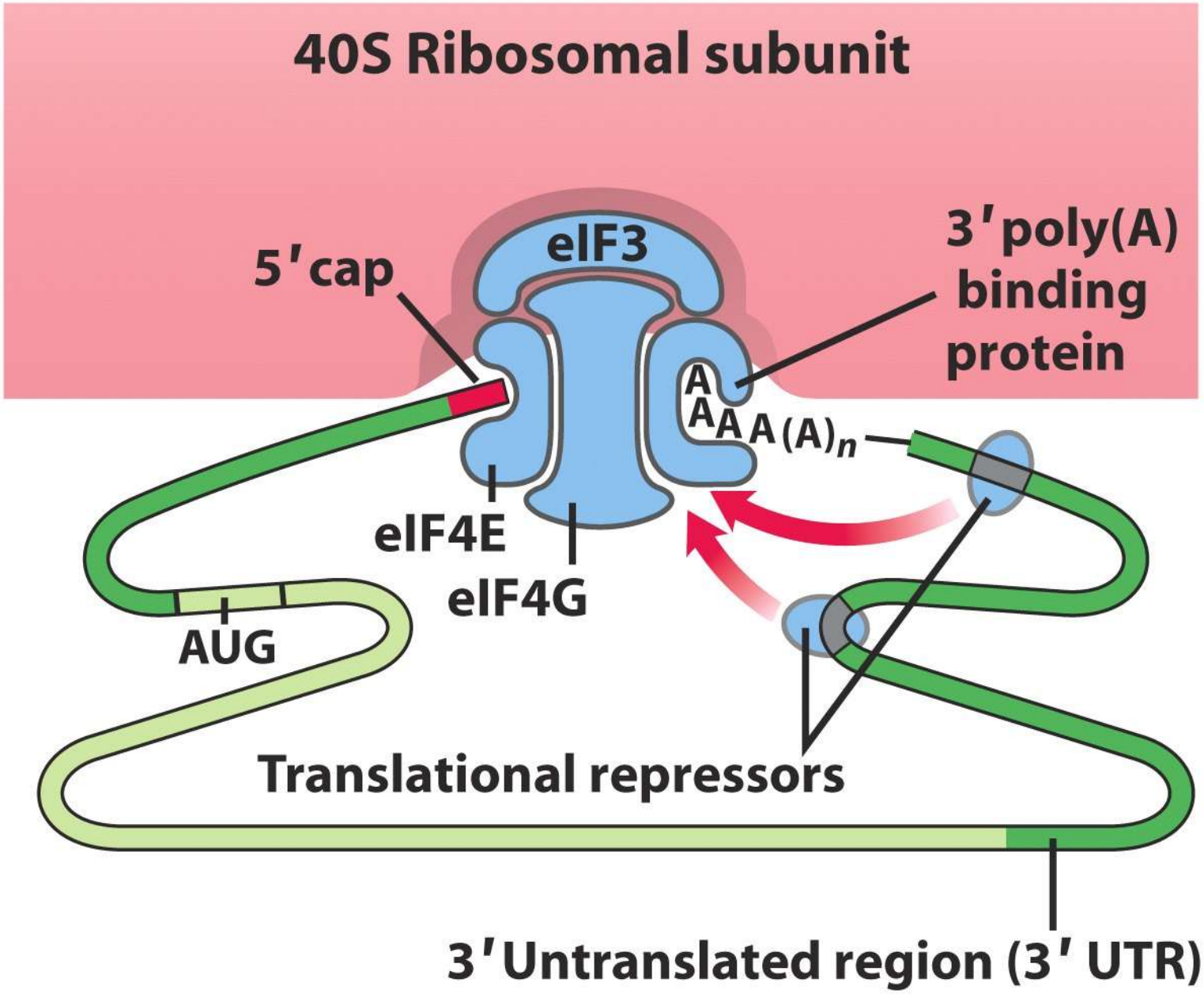
eIF4E inhibition

- eIF4E is necessary to bind the 5’ CAP in order to from an initiation complex for translation
  - Normally it binds eIF4G
- Maskin binds eIF4E (preventing it from binding eIF4G) when it is bound to an mRNA through interaction with CEPB

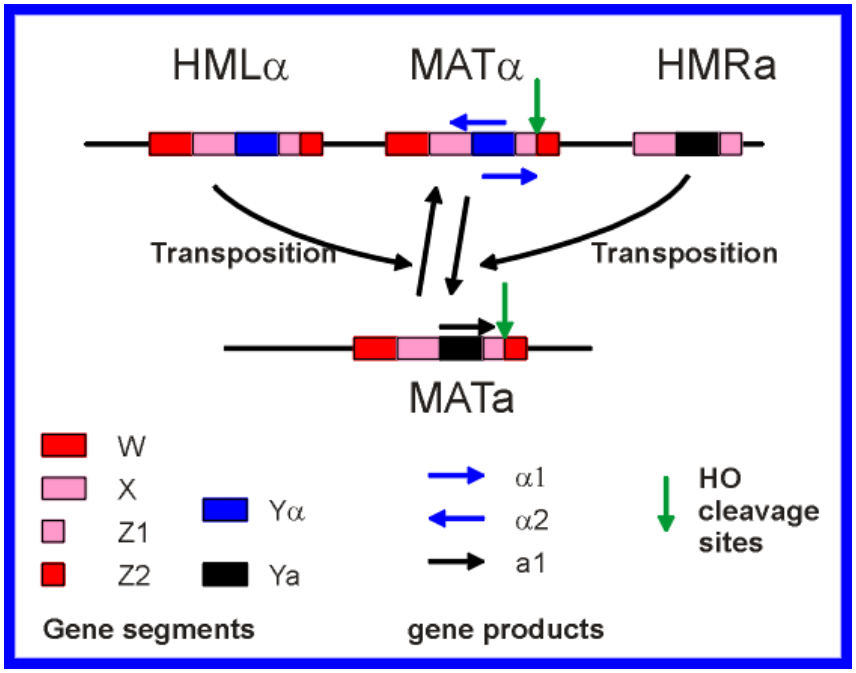


Regulation of Translation is more Common in Eukaryotes than Prokaryotes

- Phosphorylation of initiation factors can inhibit translation and cause a generalized depression of translation in cells.
  - Certain viruses phosphorylate host IFs. The viral mRNAs may not require host IFs for initiation and thus gain a selective advantage.
- Translation initiation for specific mRNAs can be inhibited by the binding of translational repressor proteins to specfic sequences in the 3' UTR.


40S Ribosomal subunit
5′cap
eIF3
3′poly(A)
binding
protein
A
AAA(A)n
eIF4E
eIF4G
AUG
Translational repressors
3′Untranslated region (3′ UTR)

Switching of mating type in yeasts



- alpha1 is a positive regulator, that switches on genes required for the alpha phenotype, including alpha factor, a secreted pheromone
- alpha2 is a negative regulator that turns off a-specific genes
- in diploid cells, a1 and alpha2 combine to inhibit alpha1 (and hence all the genes it regulates) and repress HO, and turn on the meiosis pathway if the diploid cells are starved

Regulation in Multicellular Eukaryotes

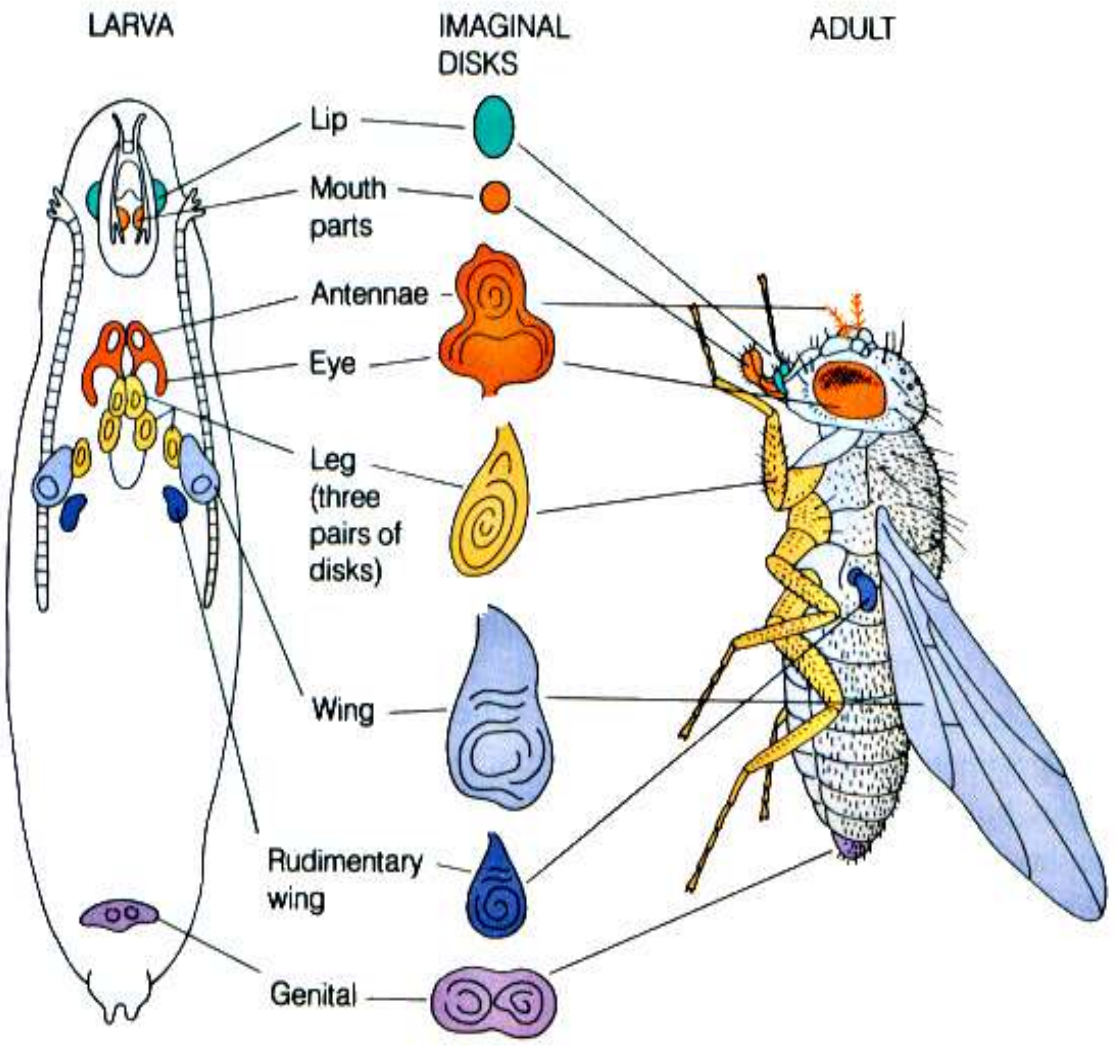
- Temporal: sequence of regulated events. Cell cycles and developmental programs are coordinated regulatory events.
- Spatial: region-specific gene expression. A 3-dimensional body plan requires specific gene activity within certain anatomical regions.
- Responsive: responses to stimuli from the environment. Just as with prokaryotes, gene expression in eukaryotes is regulated relative to the environment.

Imaginal disks in the development of *Drosophila*

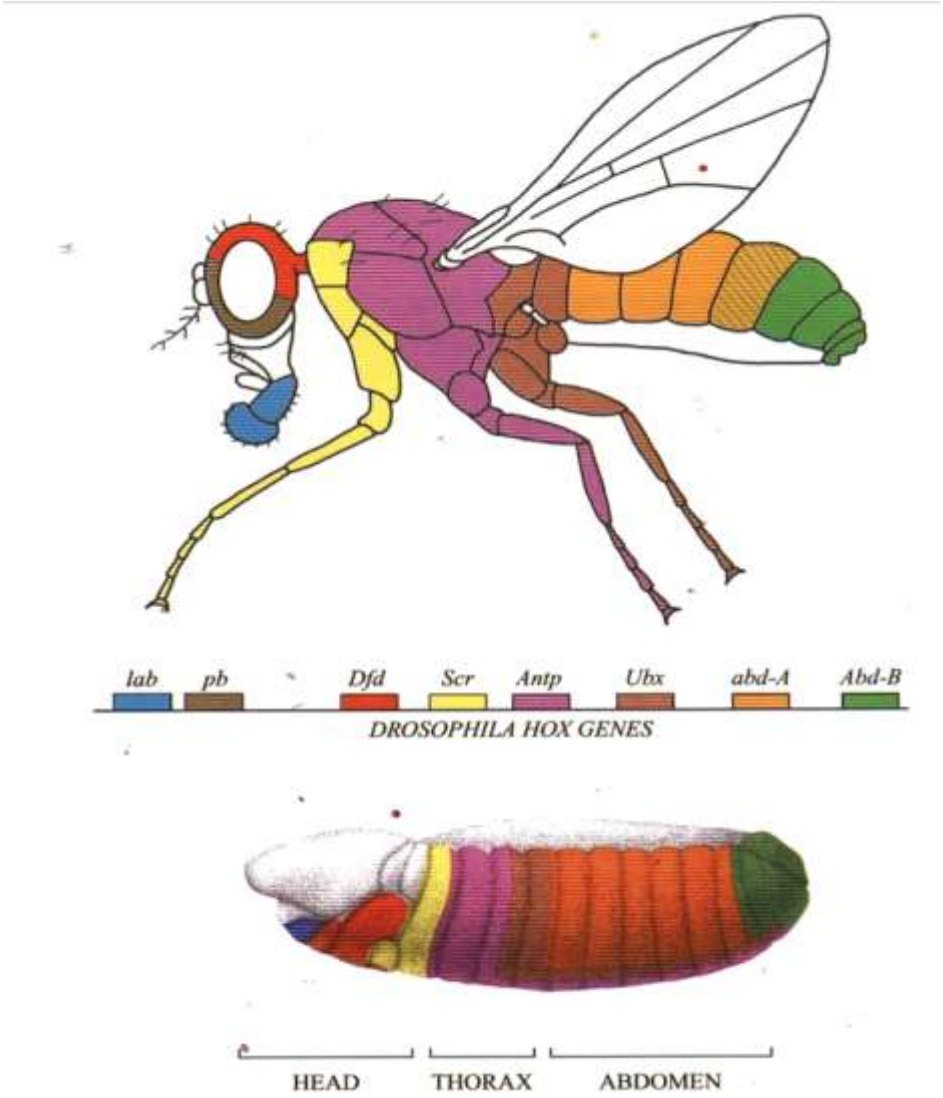
Each imaginal disk contains cells that are predetermined to develop into one particular segment of the fruit fly.

Mutations in so-called homeotic genes can redirect the development of a segment. For example, the Antennapedia mutation causes legs to grow where normally antennae grow, or the bithorax mutation leads to abnormal thorax development.

Homeotic genes code for transcription factors. They all contain a 180 bp sequence, called the homeobox which is highly conserved and found in essentially all animals.



Eight homeobox (*Hox)* genes regulate the identity of regions within the adult and embryo.

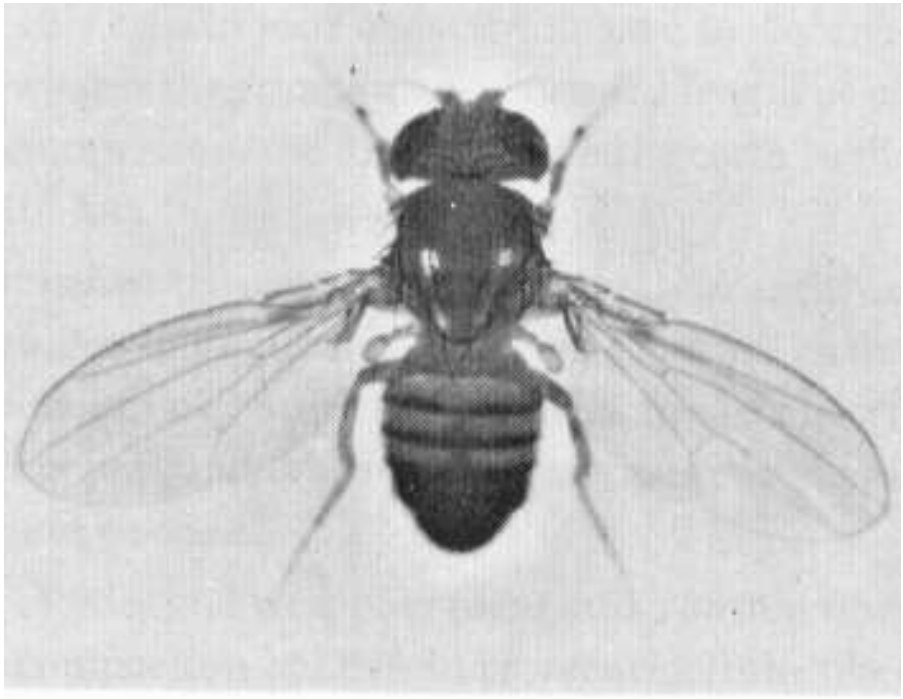


Adult

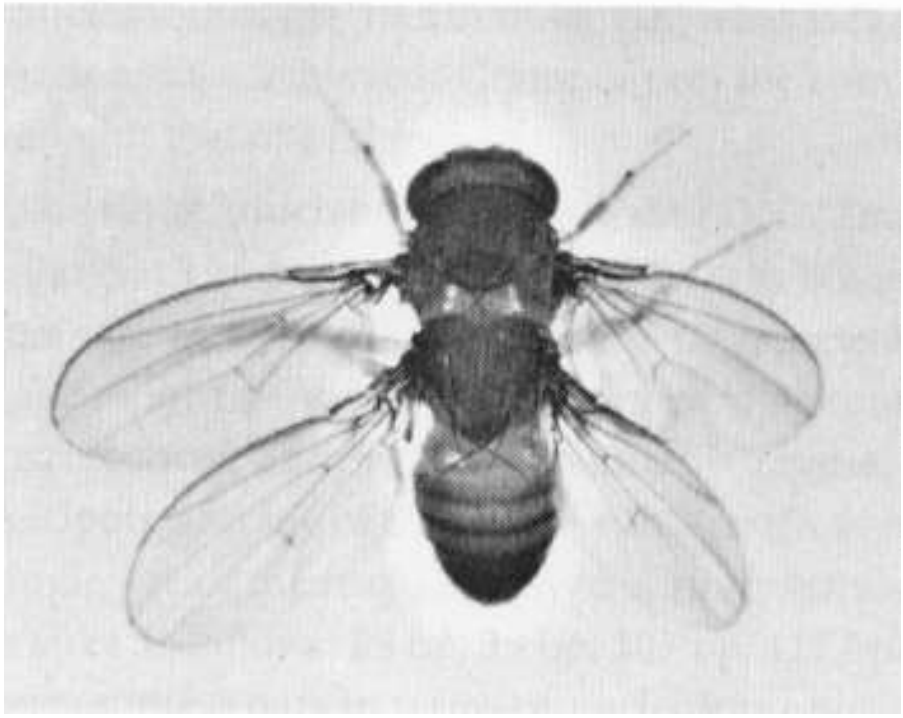
Embryo

Phenotype of a homeotic mutation:

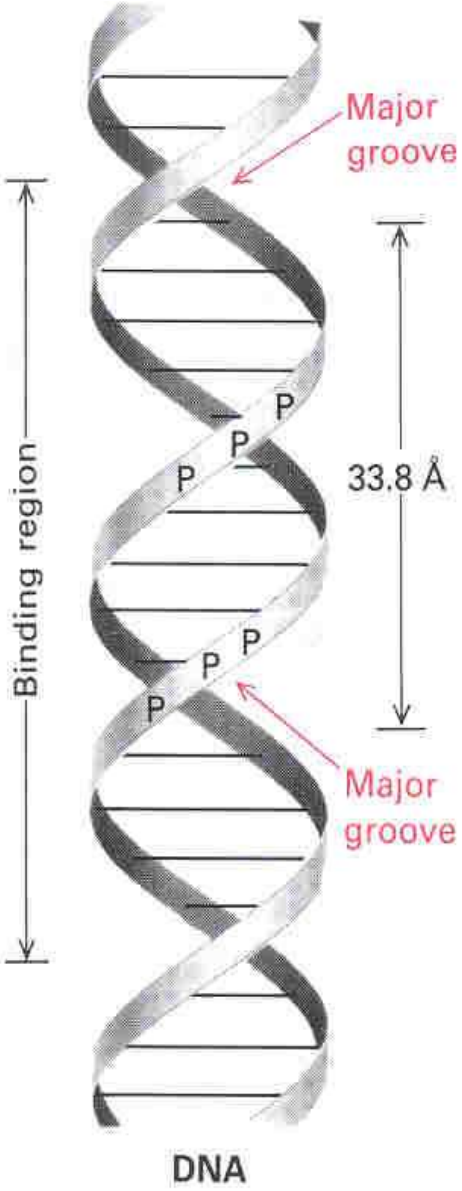
Normal *Drosophila*



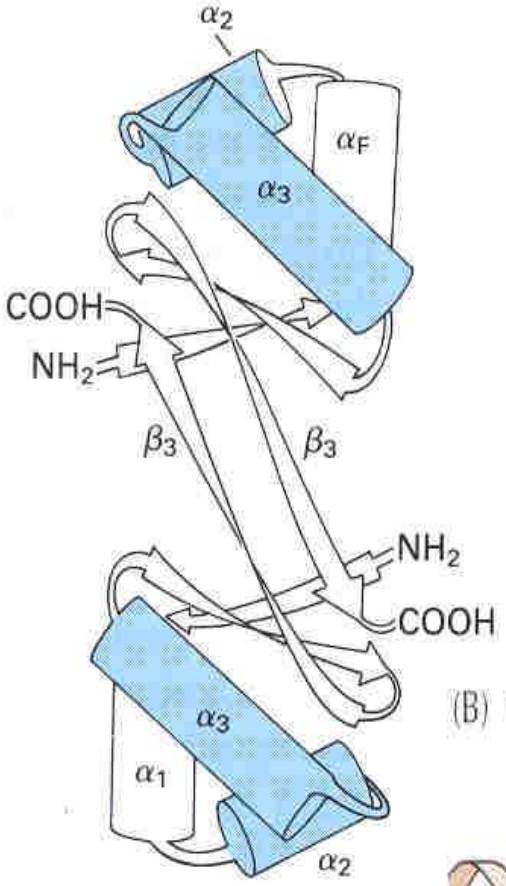
Drosophila that carries a mutation in the Ubx (ultrabithorax) homeotic gene.



Structure of homeodomain

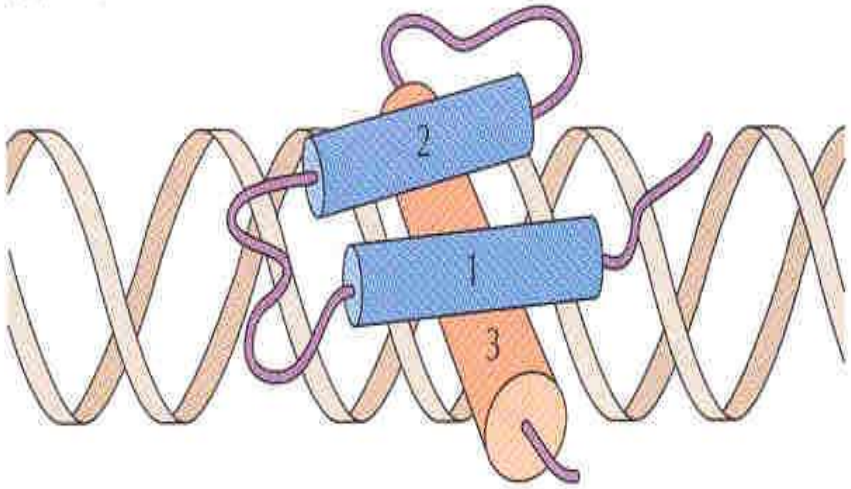


DNA



Cro protein

(B) Helix-turn-helix



Helix-Turn-Helix protein and Homeoboxes

-Three α-helical regions separated by short turns.

-Bind to DNA in similar manner.

-Bind as dimers

-Amino acid substitution in the α helices near C terminal disturb the binding in major groove of DNA

-The region encoding about 60 amino acids was remarkably conserved in all these genes called *homeobox*.

Homeobox = region in the gene

Homeotic gene = gene

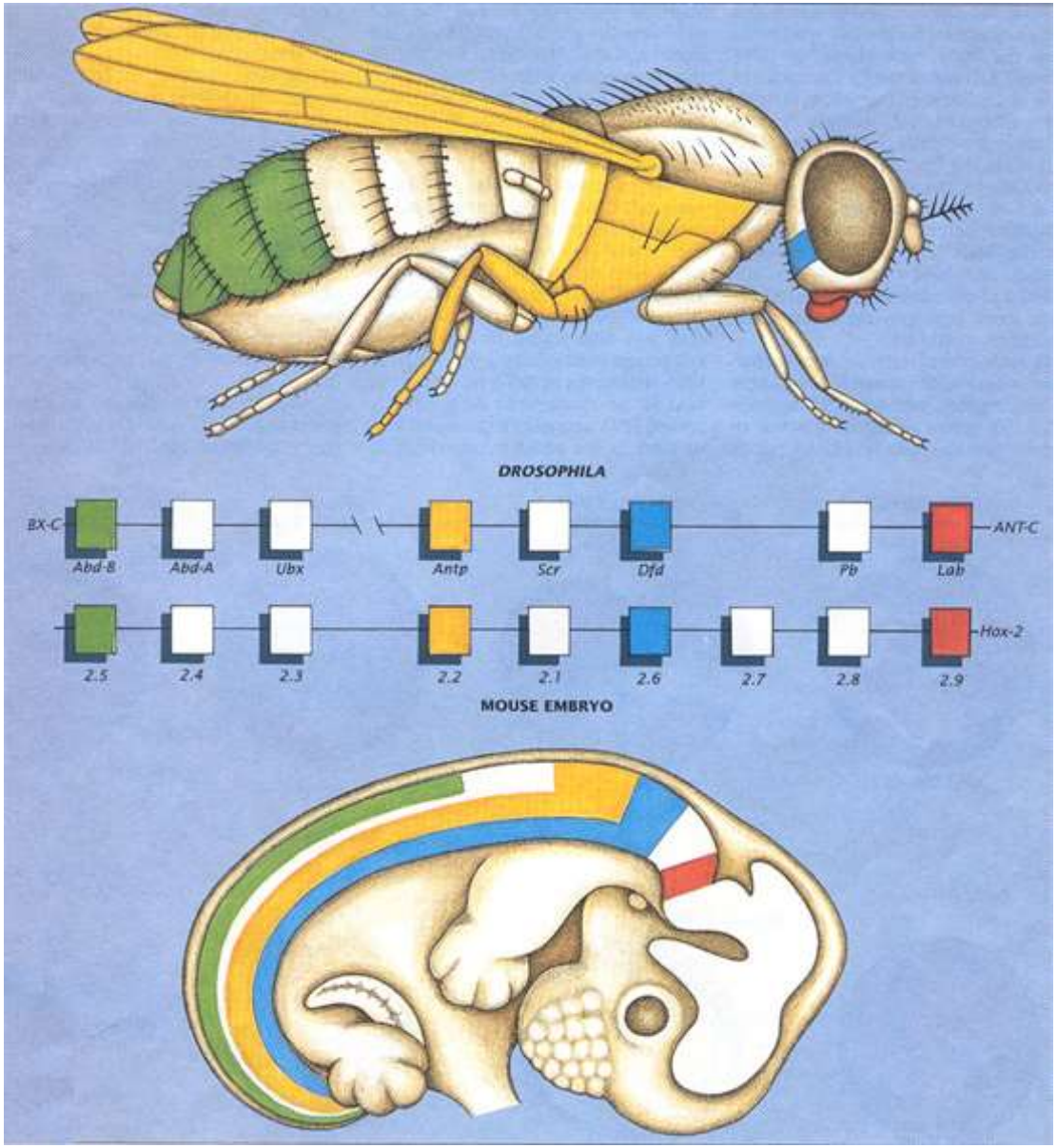
Homeodomain / helix-turn-helix protein = protein encoded by homeotic gene

The *Hox* genes are paralogs with a highly conserved homeobox domain

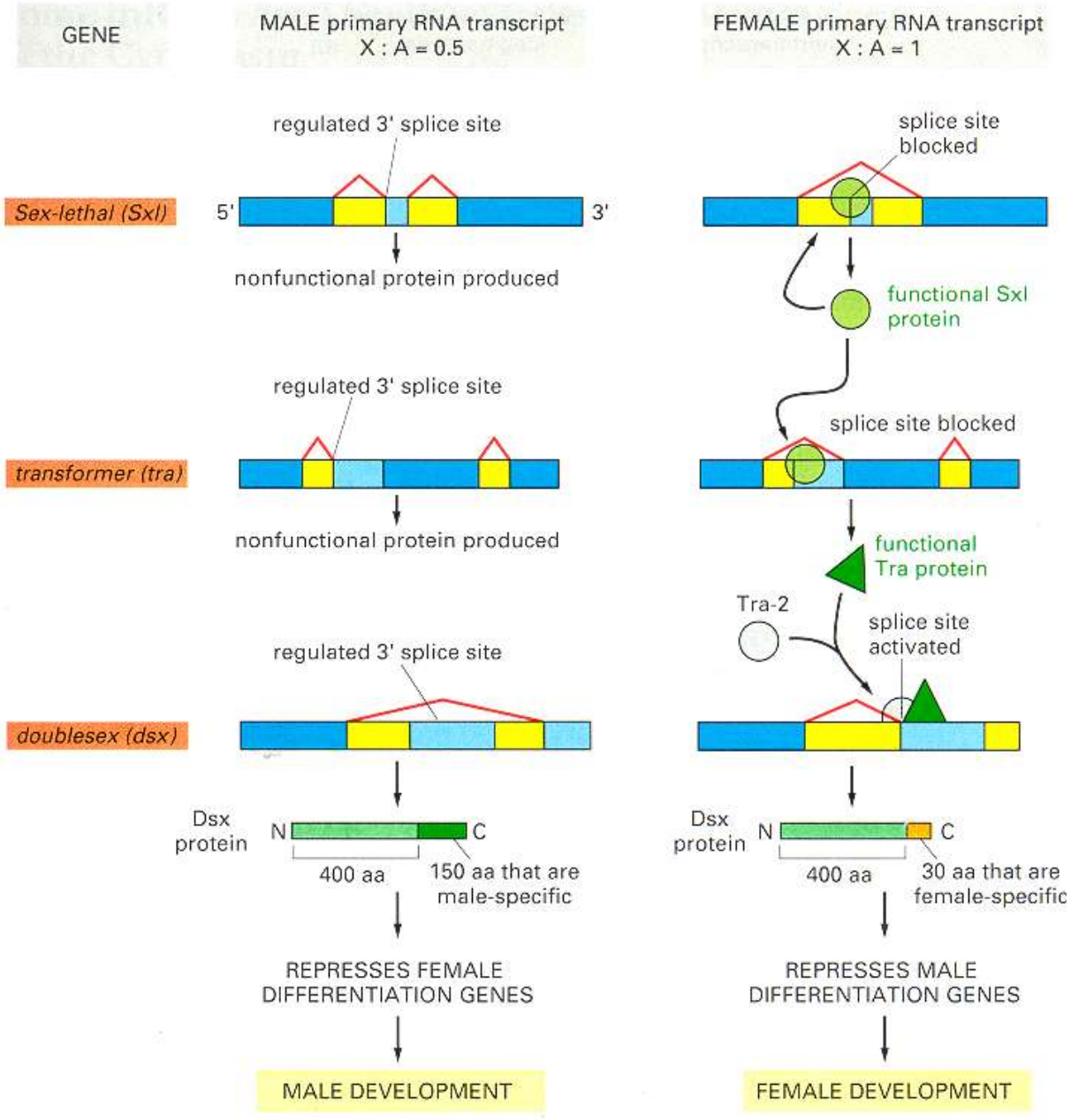
Multiple alignment of the homeobox domain:

The *Hox* genes are transcription regulators and the Homeobox domain is DNA-binding domain.

Hox complexes are conserved between *Drosophila* and mammals



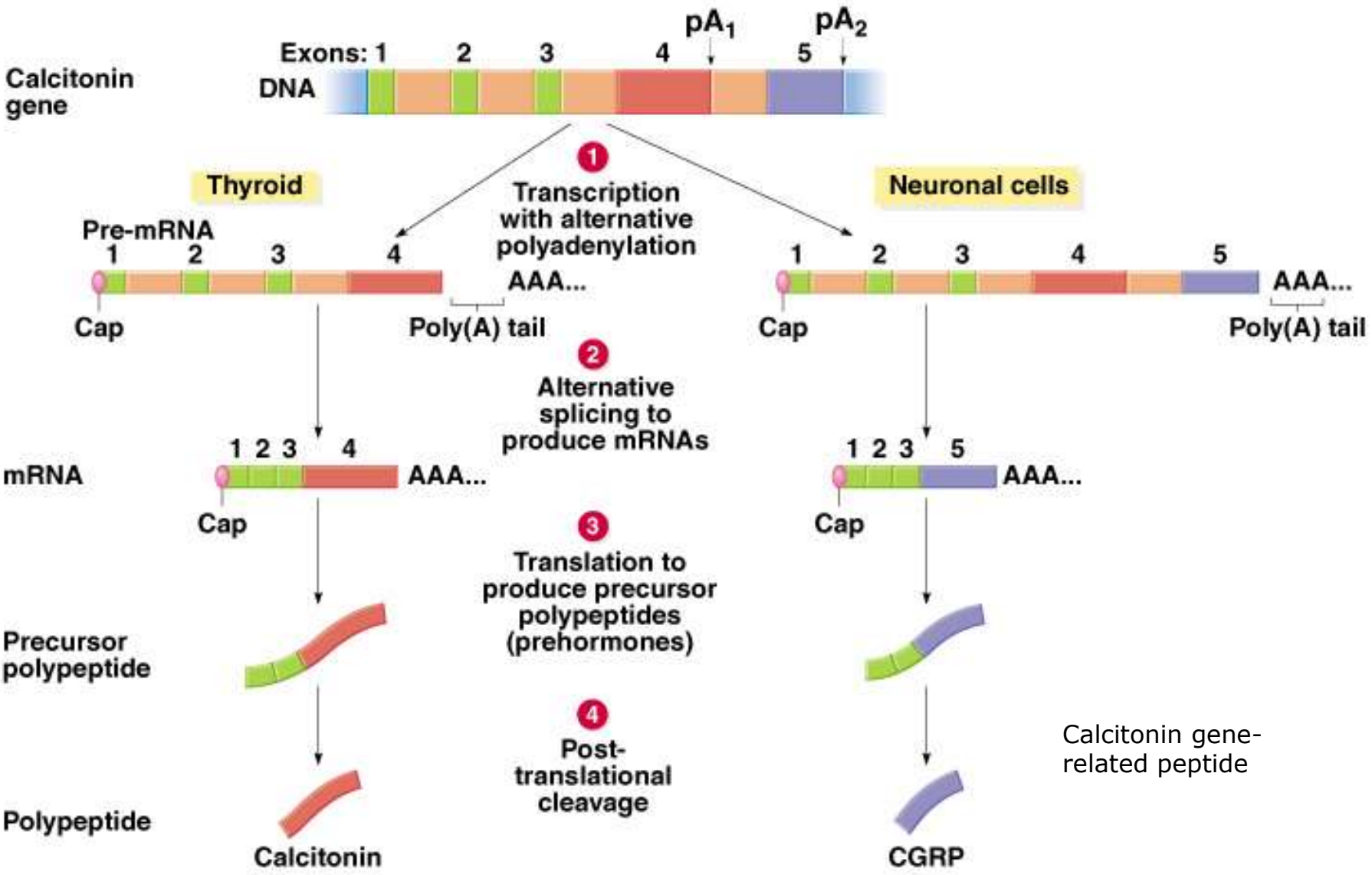
Drosophila Sex Determination



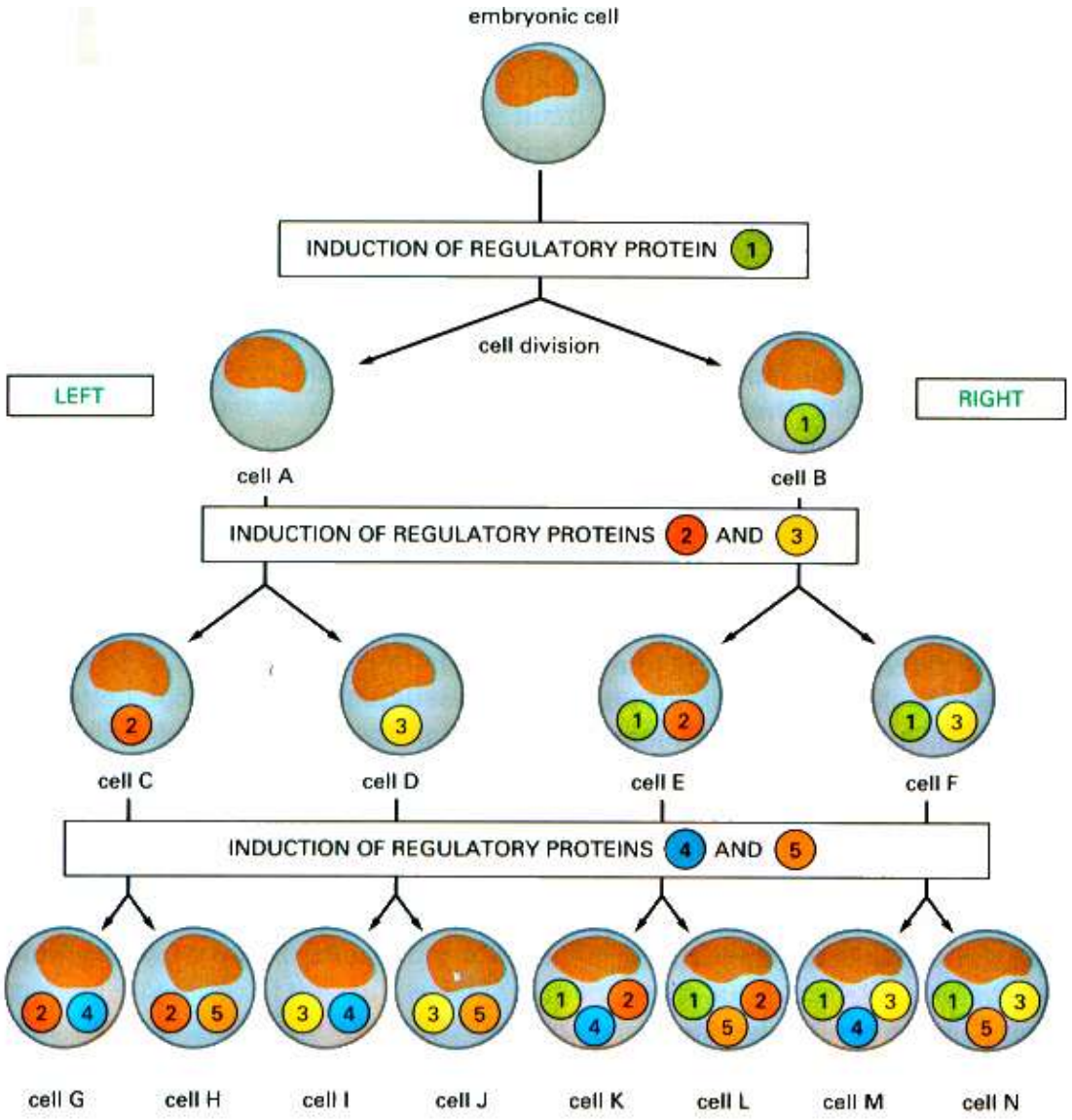
Drosophila Sex Determination

- The full SXL protein is produced in females, and absent in males, as a primary switch in the sex-determining cascade. SXL is an RNA binding protein that influences differential spicing of *sxl* and *tra* transcripts. Presence of SXL protein also suppresses the dosage compensation mechanism.
- Presence of TRA protein by SXL-mediated splicing and TRA-2 leads to female-specific splicing of *dsx*. Presence of TRA-2 (without TRA) leads to male-specific splicing of *dsx*.
- Alternative products from *dsx* control cell-specific somatic phenotypes.

Alternative polyadenylation and splicing of the human CACL gene in thyroid and neuronal cells.

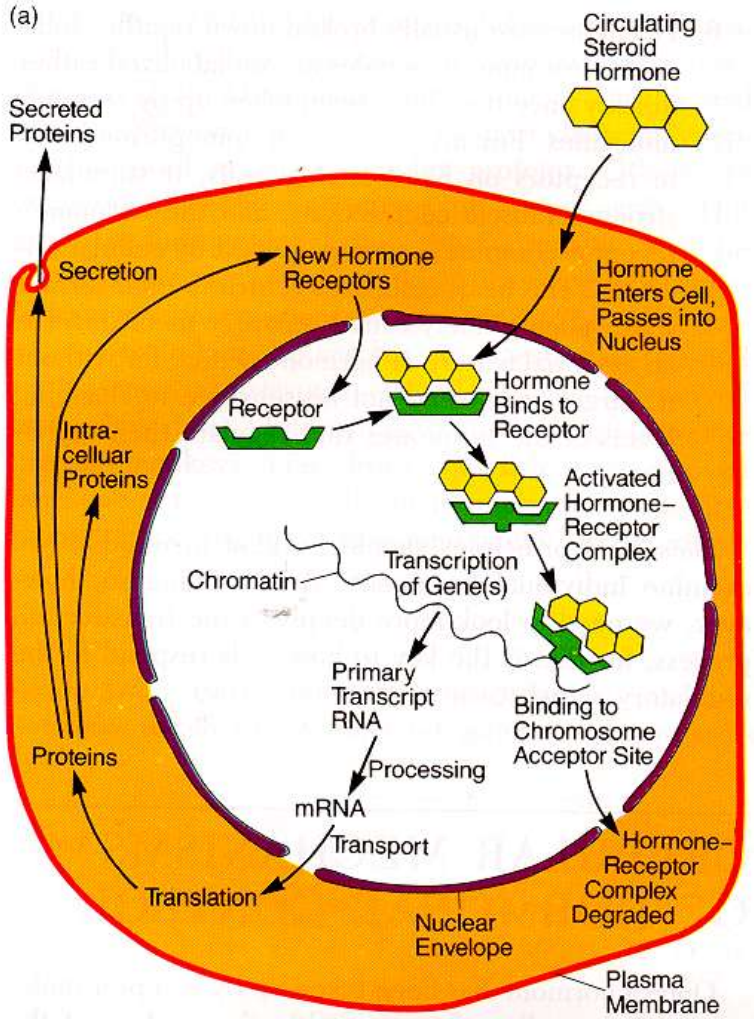


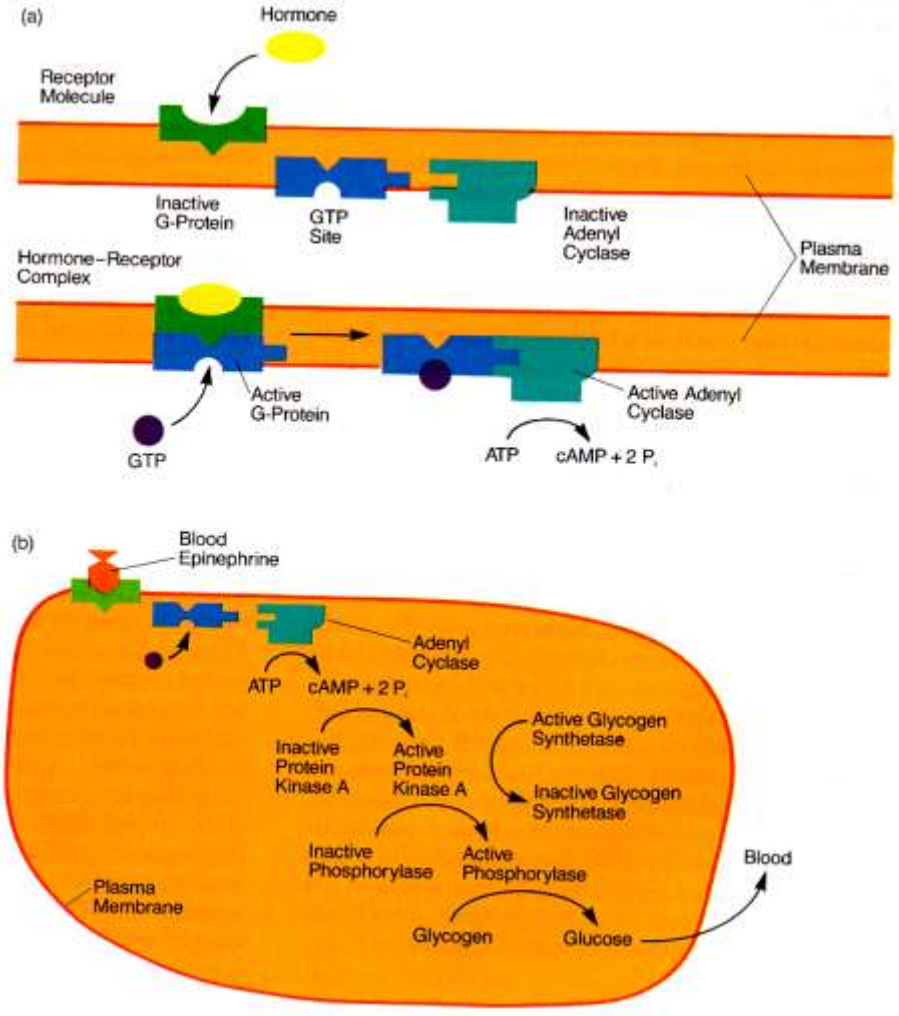
Combinatorial expression of transcription factor proteins allows differentiation of cells into different cell types



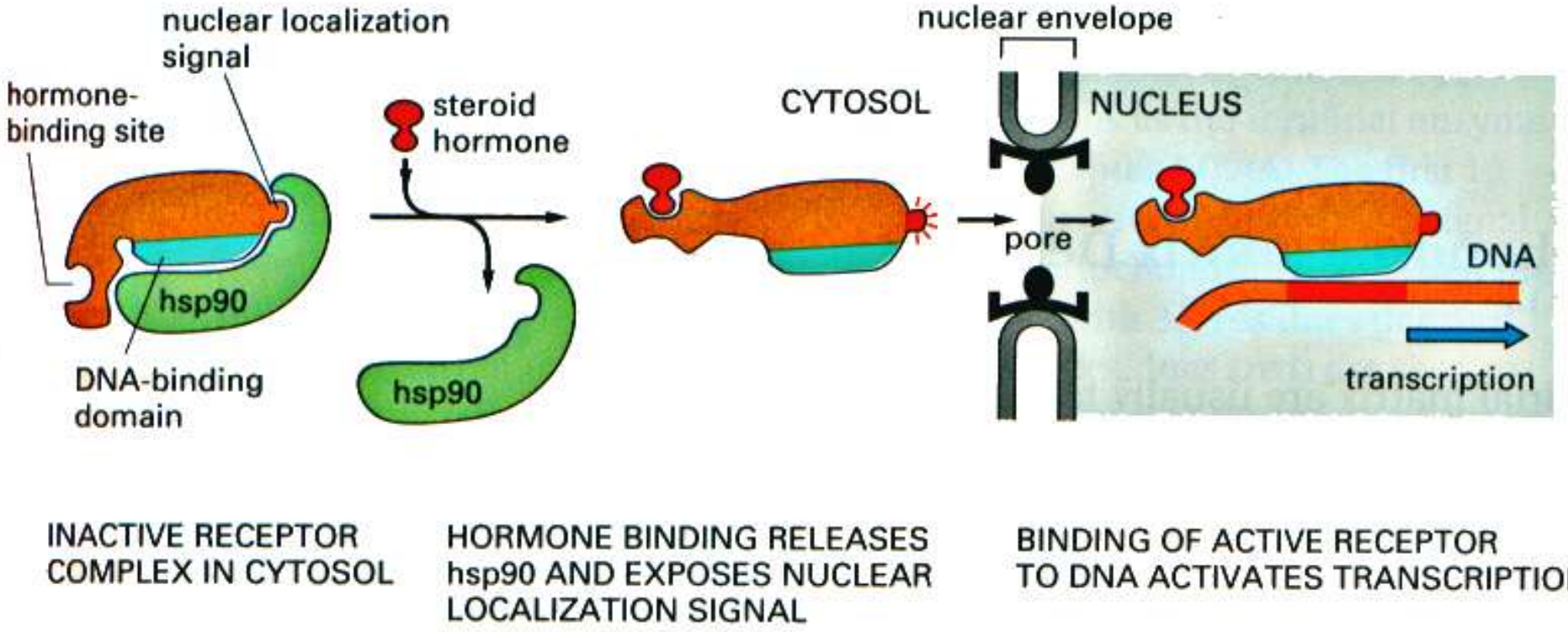
Hormones as modulators of gene expression

signal transduction





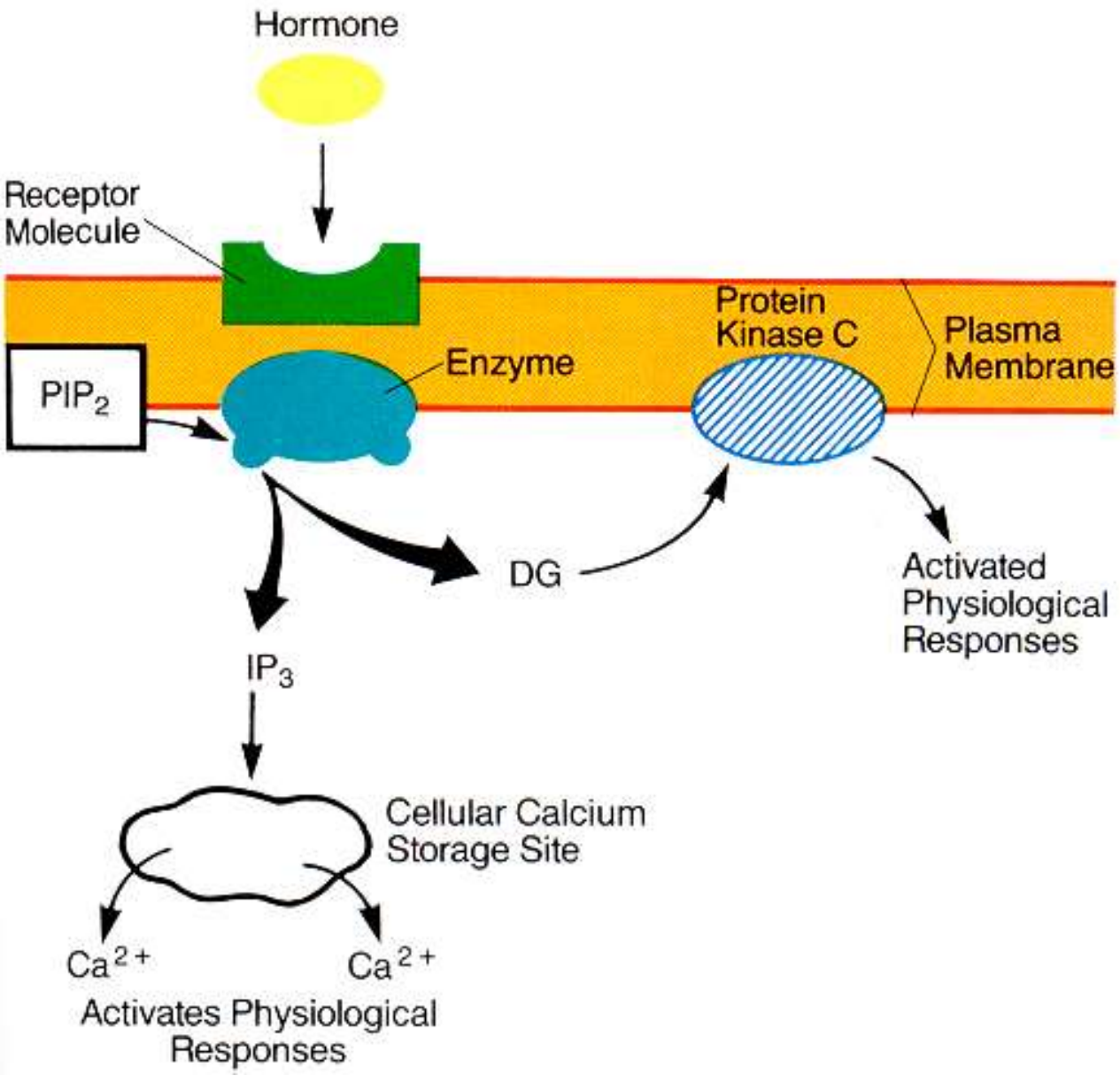
The glucocorticoid receptor is a gene regulatory protein



In the cytoplasm, the glucocorticoid receptor (GR) is bound to the chaperone protein hsp90 and in this complex is inactive. Upon binding of hormone, hsp90 dissociates and the activated GR migrates to the nucleus, where it binds to a sequence called glucocorticoid response element. This facilitates initiation of transcription.

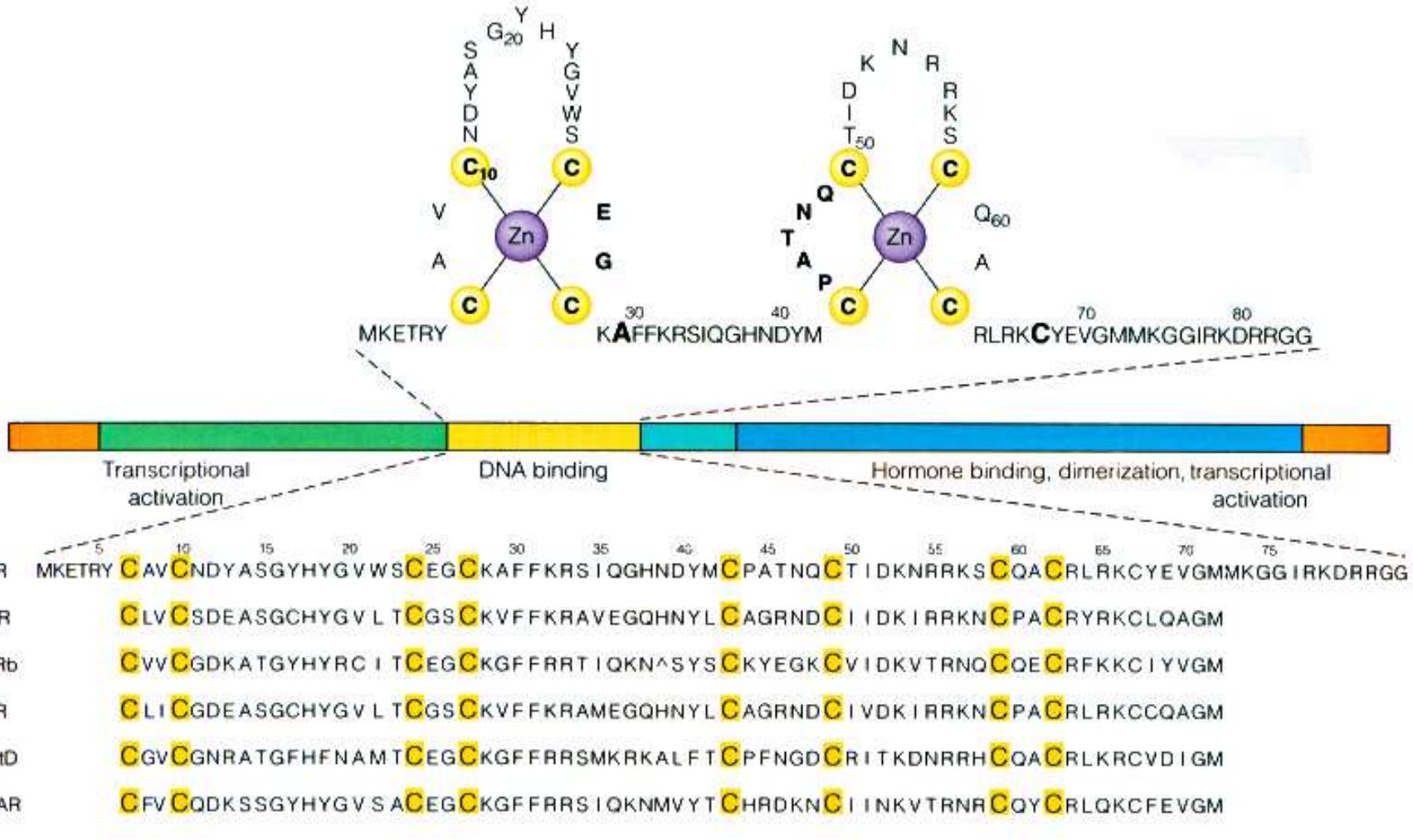
Signal Transduction

Second messenger



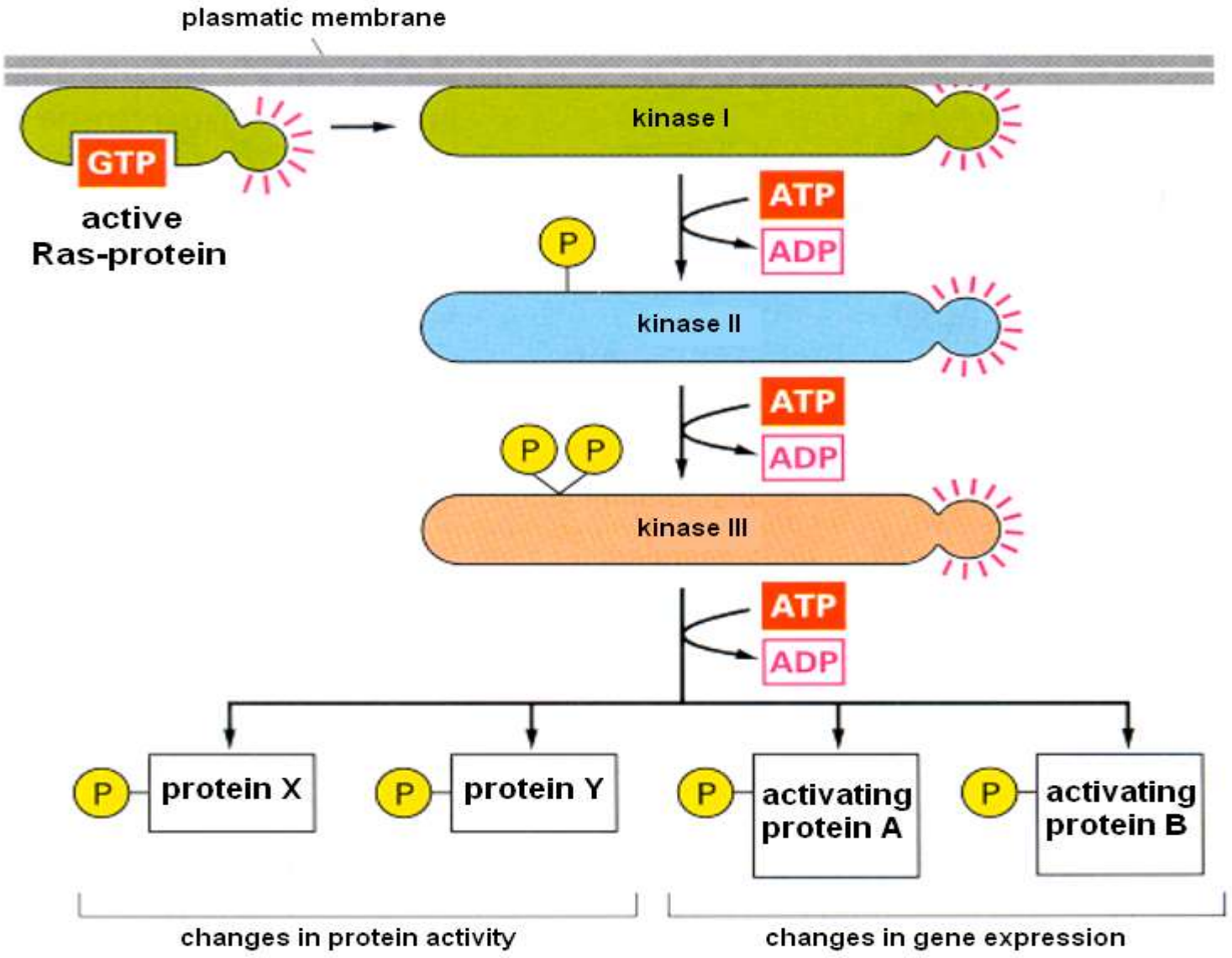
The family of steroid receptor proteins

Estrogen, glucocorticoid, thyroid hormone, progesterone, vitamin D and retinoic acid receptors share a common DNA-binding domain that forms two zinc fingers.

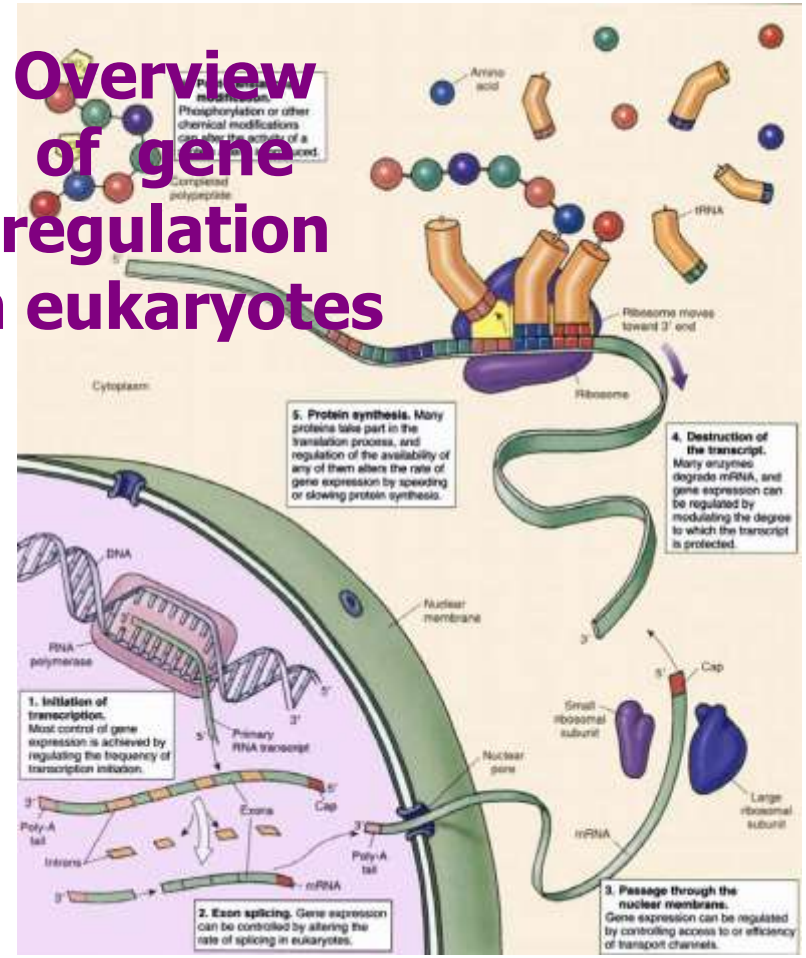


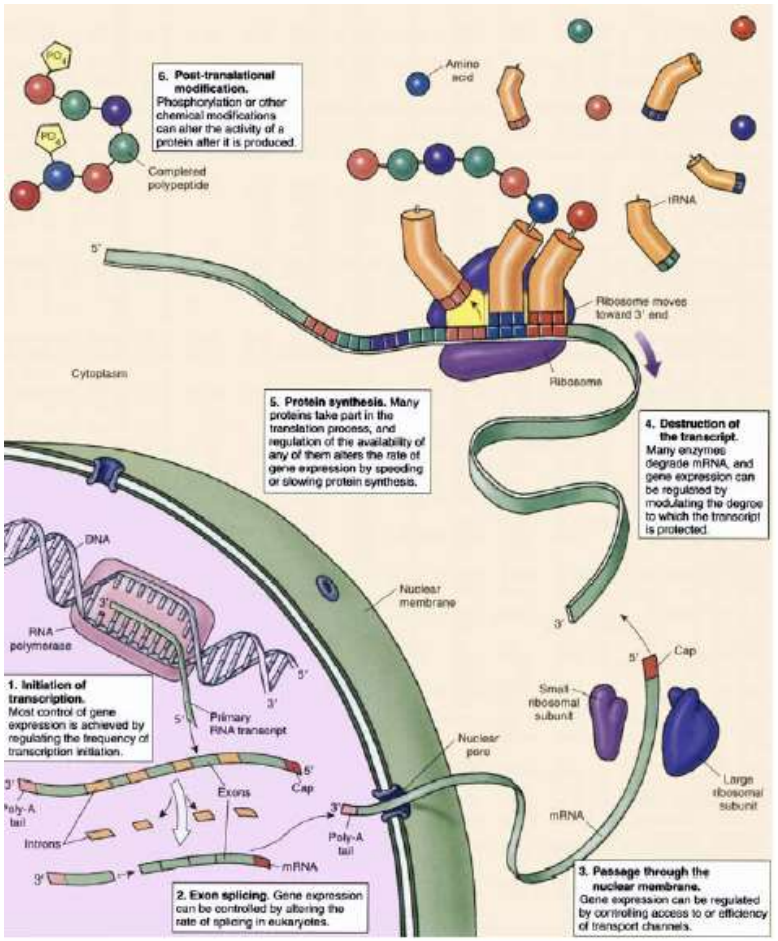
Hormones as modulators of gene expression

signal amplification



Overview of gene regulation in eukaryotes

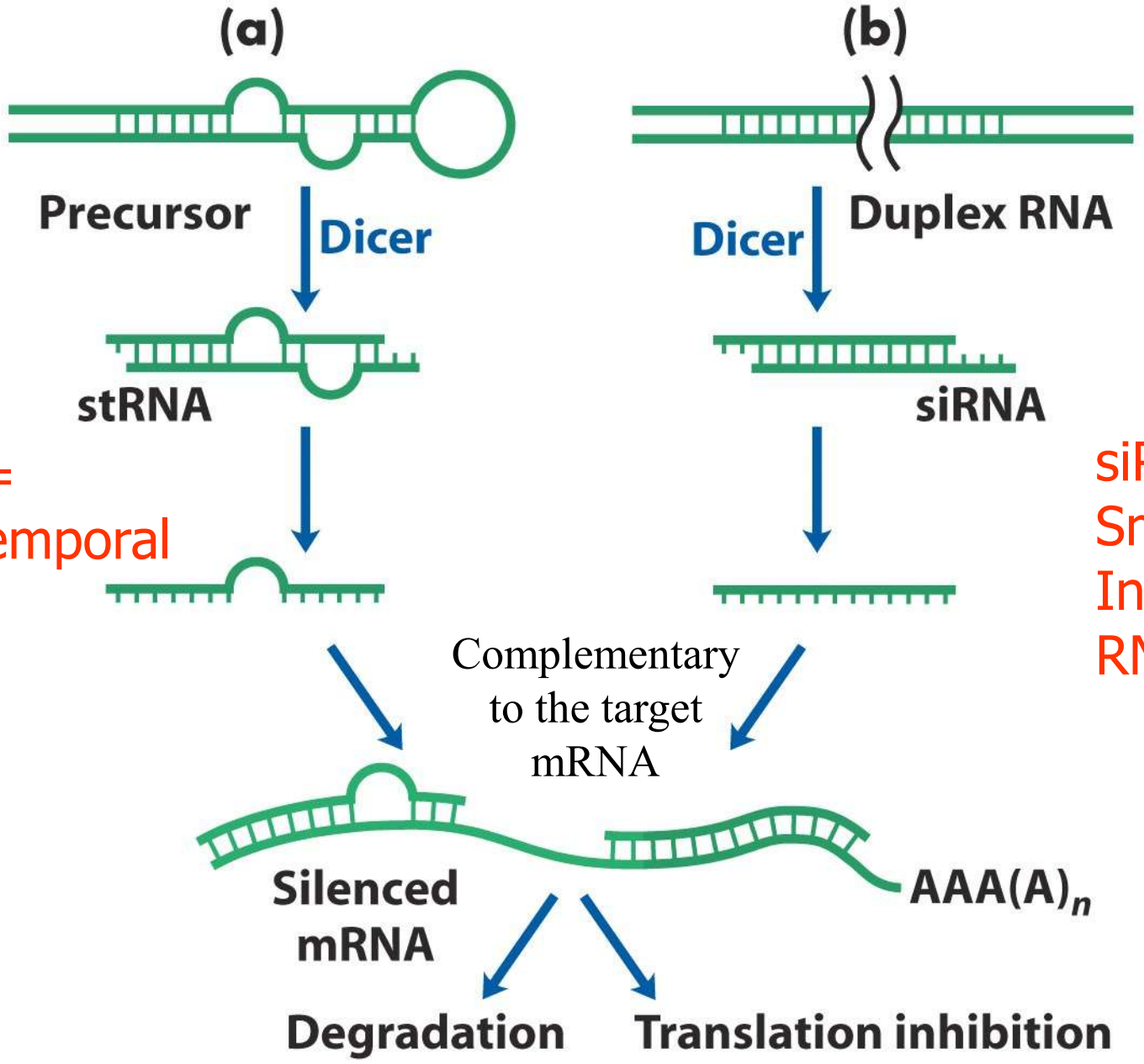




RNA-mediated Gene Silencing (RNAi = RNA interference)

- A class of small RNAs have been discovered in higher eukaryotes (worms to mammals) that shut down expression of specific genes.
- stRNAs and siRNAs interfere with their target mRNA to trigger degradation or to block translation.
- Functions in developmental gene regulation and is a defense against certain RNA viruses.
- In the lab, artifically synthesized RNAi can be used to shut down virtually any gene in plants. In worms, genes can be turned off by simply adding RNAi to the diet.
- RNAi are being developed as therapy for human diseases. Early clinical trials are aimed at:
  - blocking development of macular degeneration of the retina by RNAi silencing of a gene encoding a specific transcription factor.
  - reducing cholesterol levels in blood by RNAi silencing of apoenzyme production in the liver.

In vivo Engineered Gene Silencing



stRNA = Small temporal RNA

siRNA = Small Interfering RNA

Prokaryotes control expression by:

Transcription

Eukaryotes control expression by:

Transcription

RNA processing

mRNA transport

mRNA translation

mRNA degradation

Protein degradation