REGULATION OF TRANSCRIPTION IN EUKARYOTES

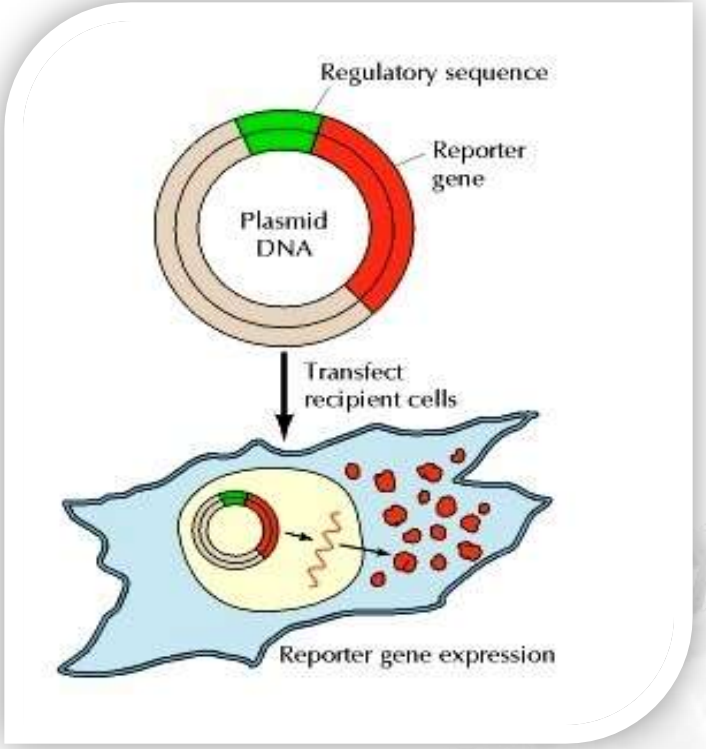
BY,
ISHAQUE P.K
BIOCHEMISTRY & MOLECULAR BIOLOGY
PONDICHERRY UNIVERSITY

INTRODUCTION

- Although the control of gene expression is far more complex in eukaryotes than in bacteria.
- The expression of eukaryotic genes is controlled primarily at the level of initiation of transcription, although in some cases transcription may be attenuated and regulated at subsequent steps.
- As in bacteria, transcription in eukaryotic cells is also controlled by proteins that bind to specific regulatory sequences and modulate the activity of RNA polymerase.
- The intricate task of regulating gene expression in the many differentiated cell types of multicellular organisms is accomplished primarily by the combined actions of multiple different transcriptional regulatory proteins.
- In addition, the packaging of DNA into chromatin and its modification by methylation impart further levels of complexity to the control of eukaryotic gene expression.

•cis-ACTING REGULATORY SEQUENCES

- Transcription in bacteria is regulated by the binding of proteins to *cis*-acting sequences (e.g., the *lac* operator) that control the transcription of adjacent genes.
- Similar *cis*-acting sequences regulate the expression of eukaryotic genes.
- These sequences have been identified in mammalian cells largely by the use of gene transfer assays to study the activity of suspected regulatory regions of cloned genes.
- Biologically active regulatory regions can thus be identified, and *in vitro* mutagenesis can be used to determine the roles of specific sequences within the region.



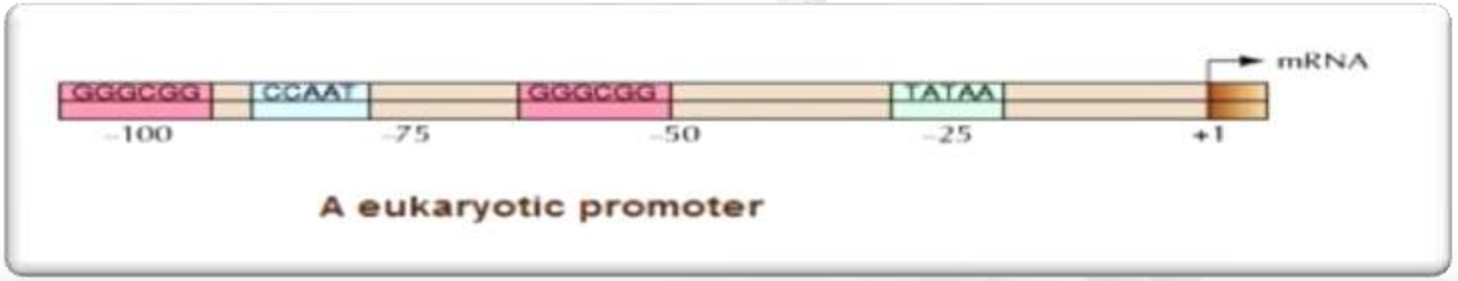
Identification of eukaryotic regulatory sequences

The regulatory sequence of a cloned eukaryotic gene is ligated to a reporter gene that encodes an easily detectable enzyme. The resulting plasmid is then introduced into cultured recipient cells by transfection. An active regulatory sequence directs transcription of the reporter gene, expression of which is then detected in the transfected cells.

•PROMOTERS AND ENHANCERS

- Genes transcribed by RNA polymerase II have two core promoter elements, the TATA box and the Inr sequence, that serve as specific binding sites for general transcription factors.
- Other cis-acting sequences serve as binding sites for a wide variety of regulatory factors that control the expression of individual genes.
- These cis-acting regulatory sequences are frequently, though not always, located upstream of the TATA box.
- For example, two regulatory sequences that are found in many eukaryotic genes were identified by studies of the promoter of the herpes simplex virus gene that encodes thymidine kinase.

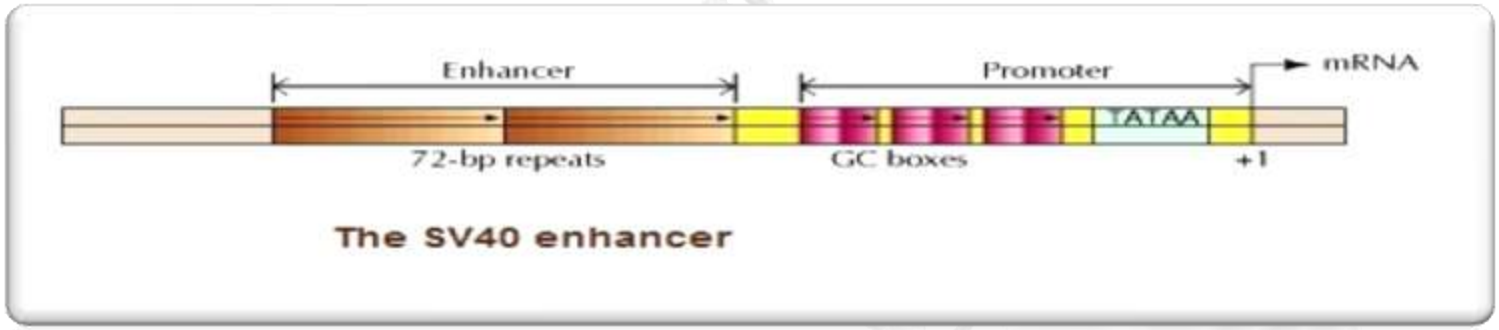
- Both of these sequences are located within 100 base pairs upstream of the TATA box: Their consensus sequences are CCAAT and GGGCGG (called a GC box).
  Specific proteins that bind to these sequences and stimulate transcription have since been identified.



A eukaryotic promoter

The promoter of the thymidine kinase gene of herpes simplex virus contains three sequence elements upstream of the TATA box that are required for efficient transcription: a CCAAT box and two GC boxes (consensus sequence GGGCGG).

- In contrast to the relatively simple organization of CCAAT and GC boxes in the herpes thymidine kinase promoter, many genes in mammalian cells are controlled by regulatory sequences located farther away (sometimes more than 10 kilobases) from thetranscription start site.
- These sequences, called enhancers, were first identified by Walter Schaffner in 1981 during studies of the promoter of another virus, SV40 .
- In addition to a TATA box and a set of six GC boxes, two 72-base-pair repeats located farther upstream are required for efficient transcription from this promoter.
- These sequences were found to stimulate transcription from other promoters as well as from that of SV40, and, surprisingly, their activity depended on neither their distance nor their orientation with respect to the transcription initiation site.

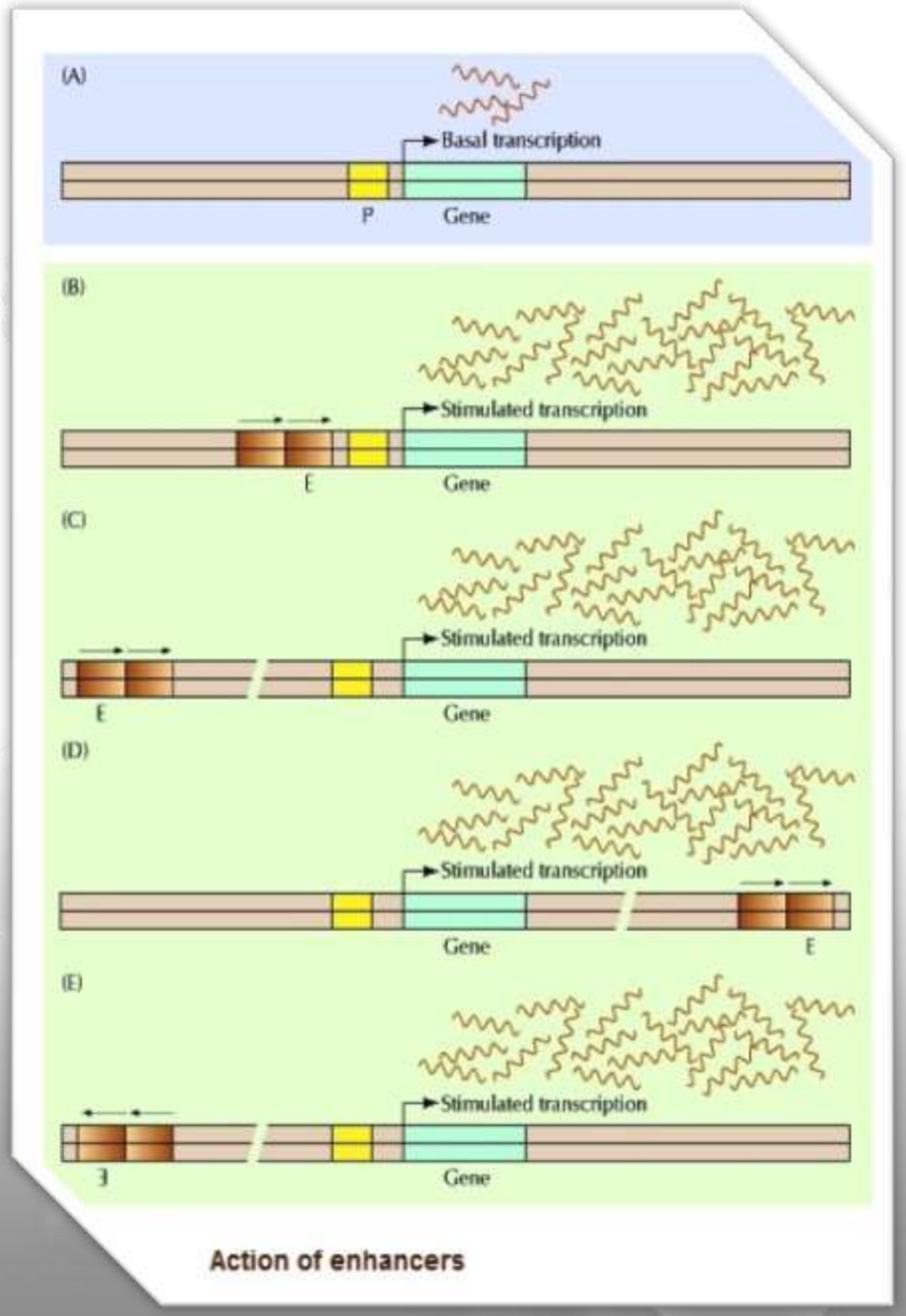


The SV40 enhancer

The SV40 promoter for early gene expression contains a TATA box and six GC boxes arranged in three sets of repeated sequences. In addition, efficient transcriptionrequires an upstream enhancer consisting of two 72-base-pair (bp) repeats.

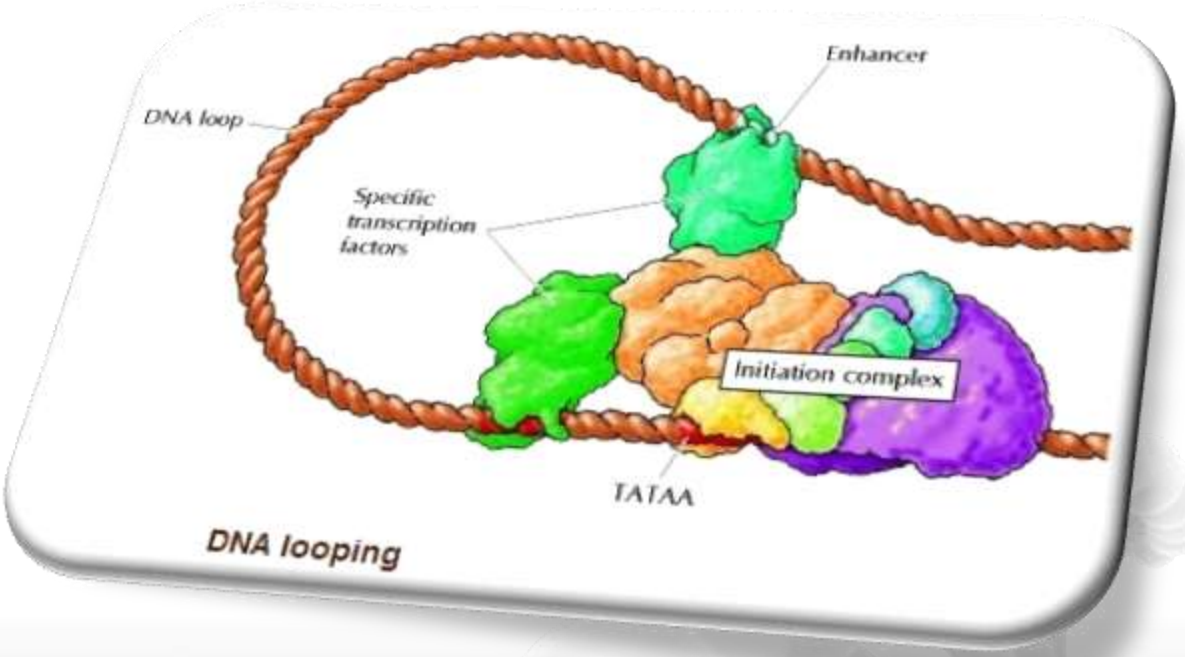
Action of enhancers

Without an enhancer, the gene is transcribed at a low basal level (A). Addition of an enhancer, E for example, the SV40 72-bp repeats stimulates the transcription. The enhancer is active not only when placed just upstream of the promotor (B), but also when inserted up to several kbp either upstream or downstream from the transcription start site (C and D). In addition, enhancers are active in either the forward or backward orientation (E).



Action of enhancers

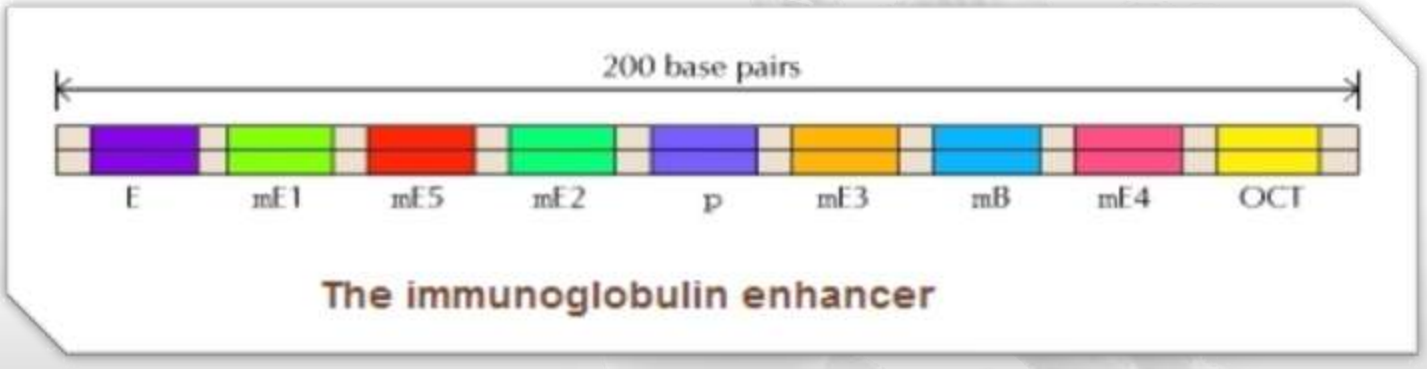
- The ability of enhancers to function even when separated by long distances from transcription initiation sites, like promoters, function by binding transcription factors that then regulate RNA polymerase.
- This is possible because of DNA looping, which allows a transcription factor bound to a distant enhancer to interact with RNA polymerase or general transcription factors at the promoter.



DNA looping

Transcription factors bound at distant enhancers are able to interact with general transcription factors at the promoter because the intervening DNA can form loops. There is therefore no fundamental difference between the action of transcription factors bound to DNA just upstream of the promoter and to distant enhancers.

The binding of specific transcriptional regulatory proteins to enhancers is responsible for the control of gene expression during development and differentiation, the immunoglobulin enhancer is active in lymphocytes, but not in other types of cells. Thus, this regulatory sequence is at least partly responsible for tissue-specific expression of the immunoglobulin genes in the appropriate differentiated cell type.



The immunoglobulin enhancer

The immunoglobulin enhancer

The immunoglobulin heavy-chain enhancer spans about 200 bases and contains nine functional sequence elements (E, μE1-5, π, μB, and OCT), which together stimulatetranscription in B lymphocytes.

Features Of Cis-acting Elements

(1) Promoter

- Core promoter
  - in eukaryote: TATA-box, Initiator (Inr).
  - in prokaryote: -10 region, Inr.
- Proximal elements of promoter
  - in prokaryote: -35 region.
  - in eukaryote: CAAT-box, GC-box.

UPE: upstream promoter element.

UAS: upstream activating sequence.

(2) Terminator

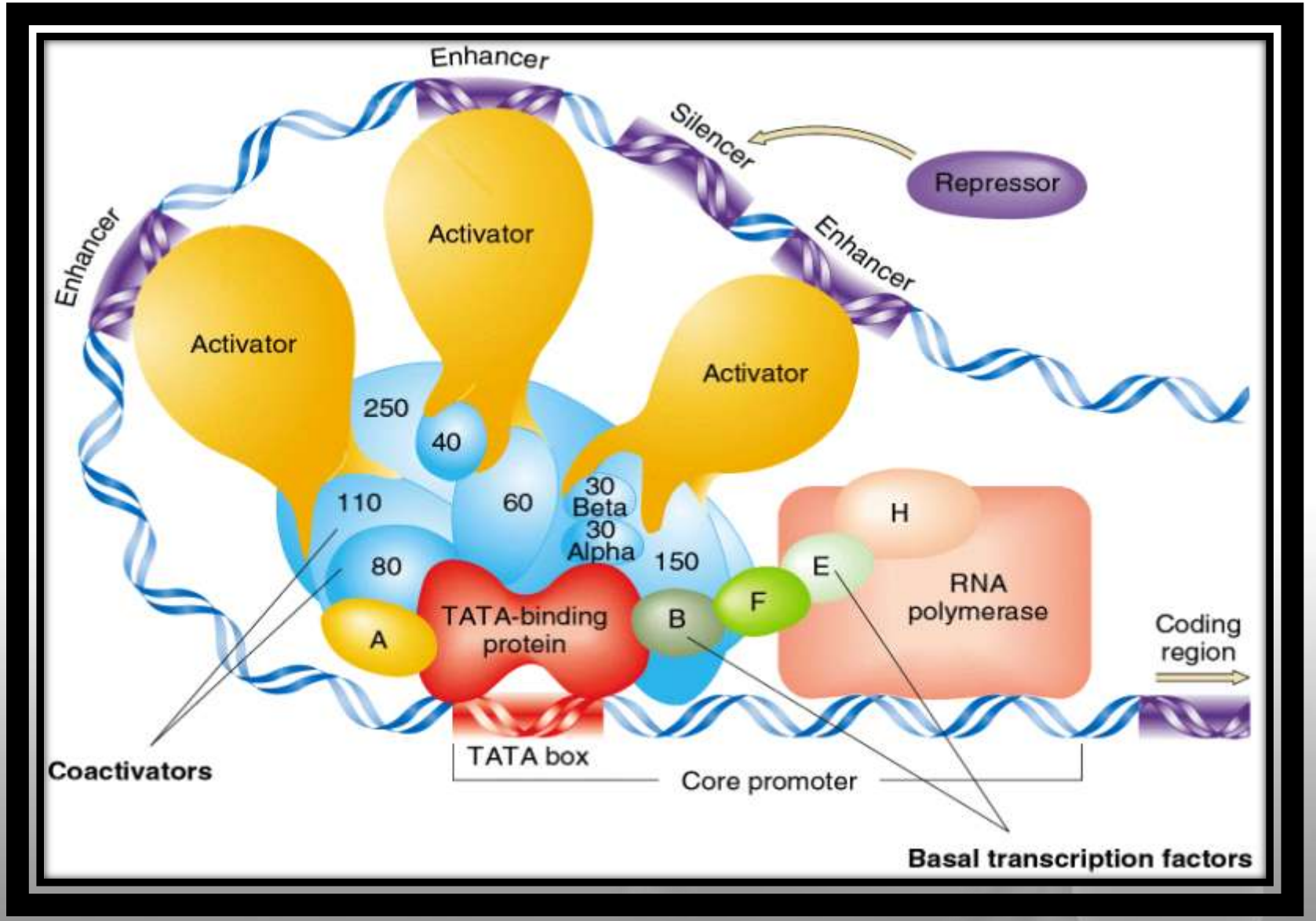
A DNA sequence just downstream of the coding segment of a gene, which is recognized by RNA polymerase as a signal to stop transcription.

(3) Enhancer

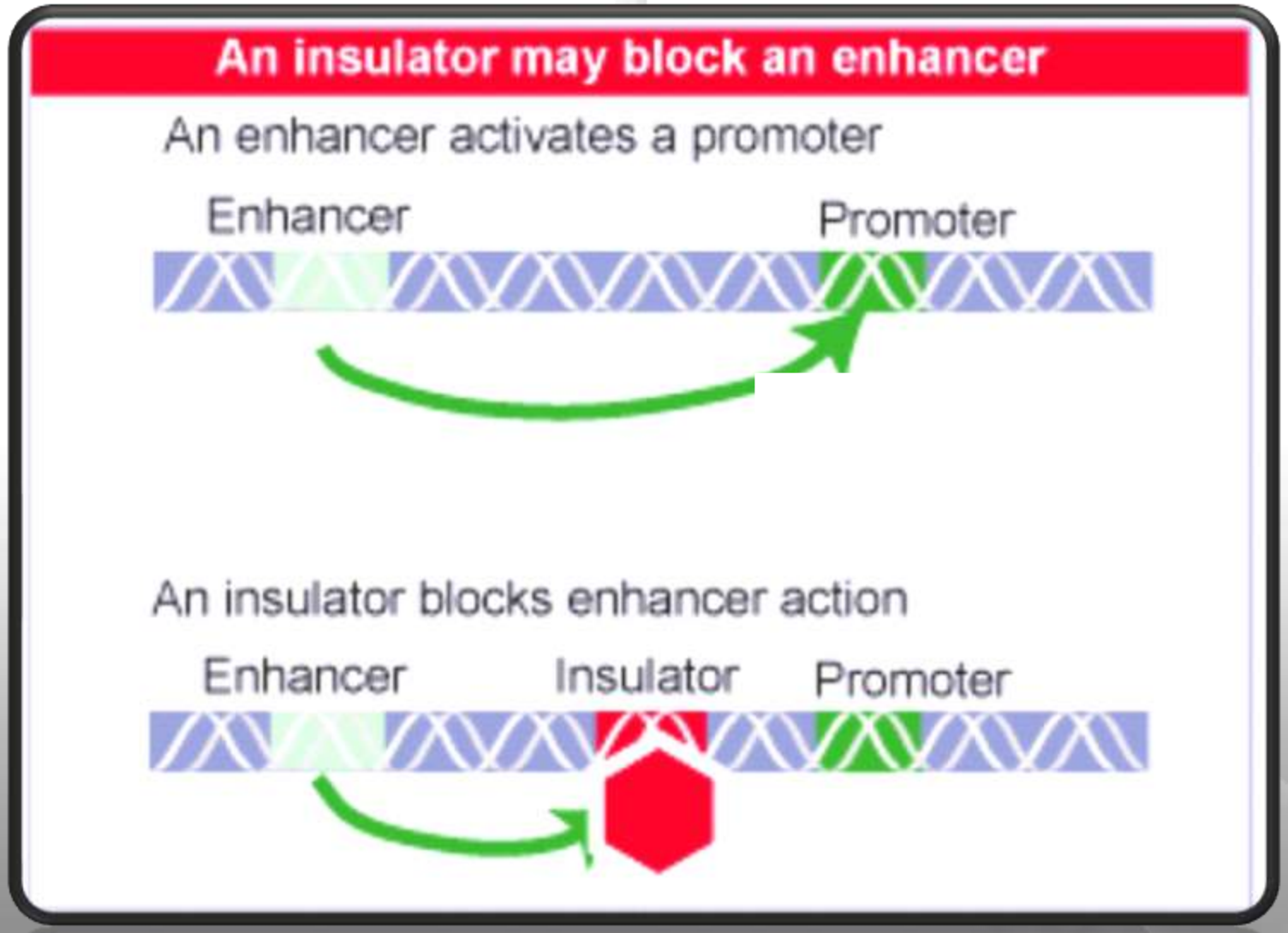
A regulatory DNA sequence that greatly enhances the transcription of a gene.

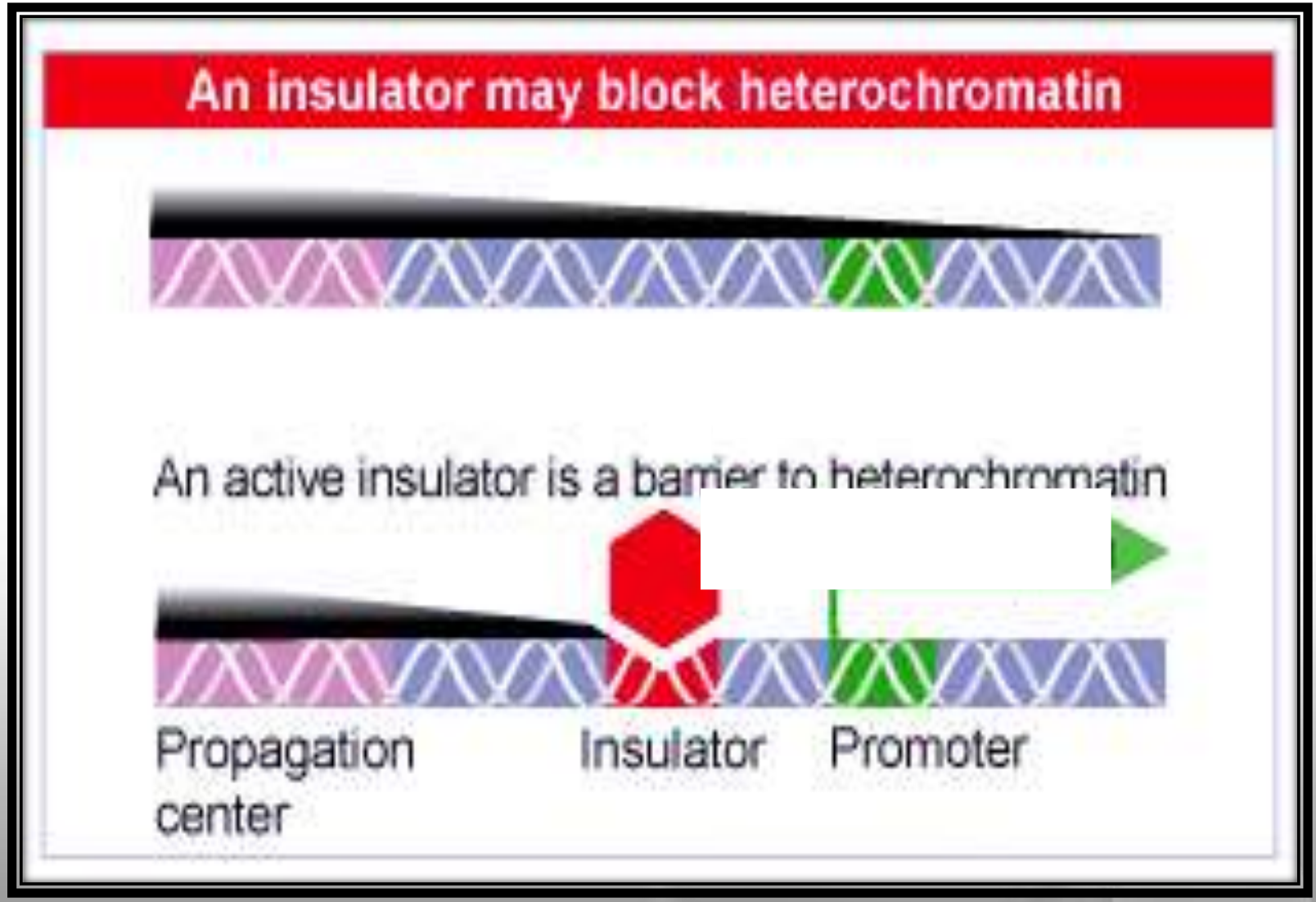
(4) Silencer

A DNA sequence that helps to reduce or shut off the expression of a nearby gene.


Enhancer
Silencer
Repressor
Enhancer
Enhancer
Activator
Activator
Activator
250
40
110
60
30
Beta
30
Alpha
150
80
H
E
RNA
polymerase
F
B
A
TATA-binding
protein
Coding
region
Coactivators
TATA box
Core promoter
Basal transcription factors

(5) Insulators




An insulator may block heterochromatin
An active insulator is a barrier to heterochromatin
Propagation center
Insulator
Promoter

•Transcriptional Regulatory Proteins

- One of the prototypes of eukaryotic transcription factors was initially identified by Robert Tjian and his colleagues during studies of the transcription of SV40 DNA.
- This factor (called Sp1, for specificity protein 1) was found to stimulate transcription from the SV40promoter, but not from several other promoters, in cell-free extracts.
- Then, stimulation of transcription by Sp1 was found to depend on the presence of the GC boxes in the SV40 promoter: If these sequences were deleted, stimulation by Sp1 was abolished.
- Moreover, footprinting experiments established that Sp1 binds specifically to the GC box sequences.

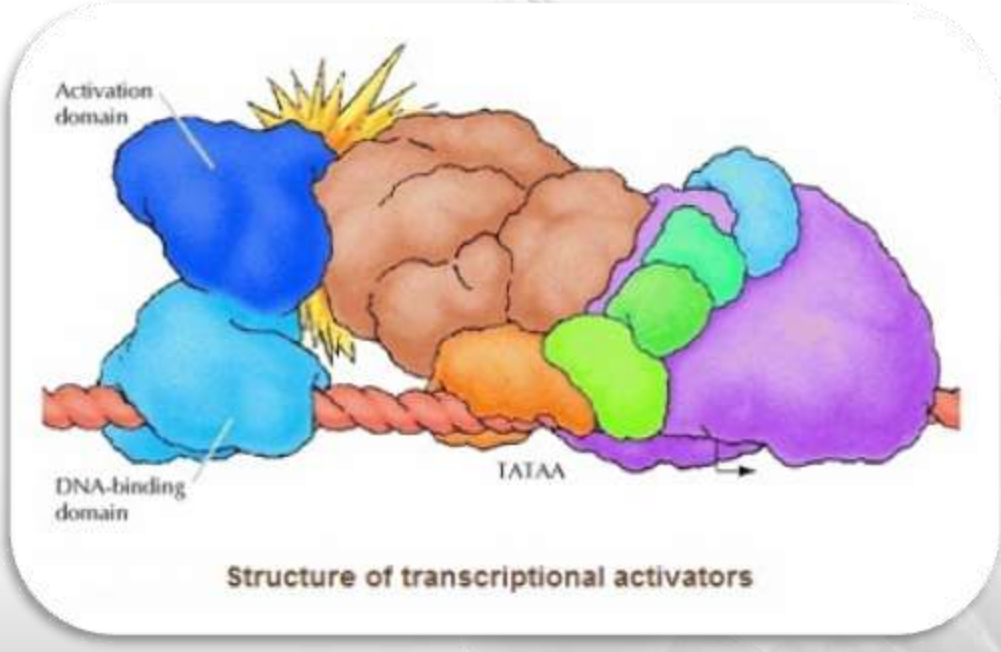
Examples of Transcription Factors and Their DNA-Binding Sites

Transcription factor	Consensus binding site
Specificity protein 1 (Sp1)	GGGCGG
CCAAT/Enhancer binding protein (C/EBP)	CCAAT
Activator protein 1 (AP1)	TGACTCA
Octamer binding proteins	ATGCAAAT
(OCT-1 and OCT-2)	
E-box binding proteins (E12, E47, E2-2)	CANNTG[a]

a N stands for any nucleotide.

The isolation of a variety of transcriptional regulatory proteins has been based on their specific binding to promoter or enhancer sequences. Protein binding to these DNA sequences is commonly analyzed by two types of experiments. The first, **footprinting**. The second approach is the **electrophoretic-mobility shift assay**, in which a radiolabeled DNA fragment is incubated with a protein preparation and then subjected to electrophoresis through a non denaturing gel. The general approach of **DNA-affinity chromatography**, has been used successfully to isolate a wide variety of sequence-specific DNA-binding protein from eukaryotic cells

•Structure and Function of Transcriptional Activators

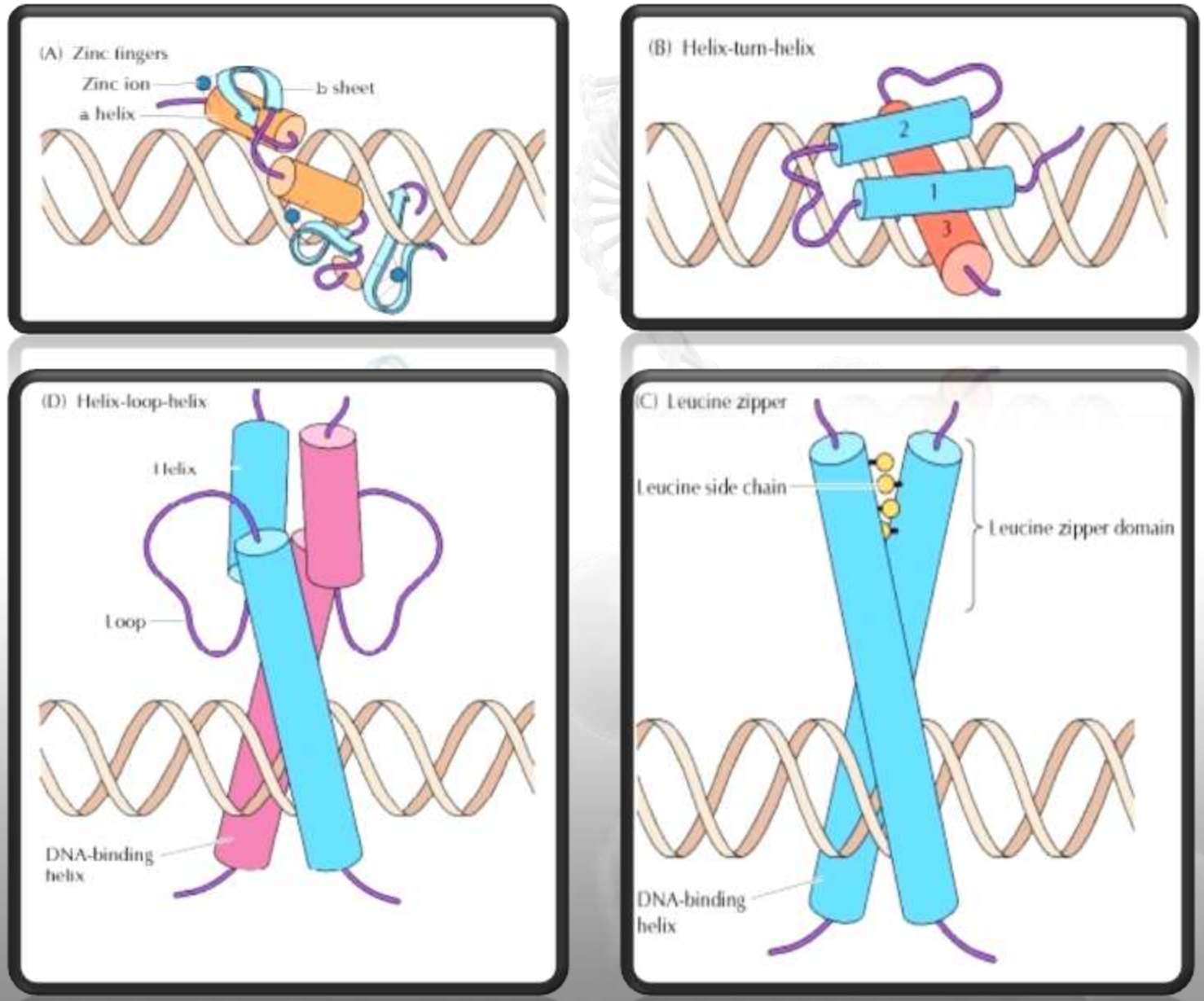


Structure of transcriptional activators

Structure of transcriptional activators consist of two independent domains. The **DNA-binding domain** recognizes a specific DNA sequence, and the **activation domain** interacts with other components of the transcriptional machinery

- The basic function of the DNA-binding domain is to anchor the transcription factor to the proper site on DNA; the activation domain then independently stimulates transcription by interacting with other proteins.
- Molecular characterization has revealed that the DNA-binding domains of many of these proteins are related to one another .
- **Zinc finger domains** contain repeats of cysteine and histidine residues that bind zinc ions and fold into looped structures ("fingers") that bind DNA.
- These domains were identified in the polymerase III transcription factor TFIIIA but are also common among transcription factors that regulate polymerase II promoters, including Sp1.
- Other examples of transcription factors that contain zinc finger domains are the steroid hormone receptors, which regulate gene transcription in response to hormones such as **estrogen** and **testosterone**.

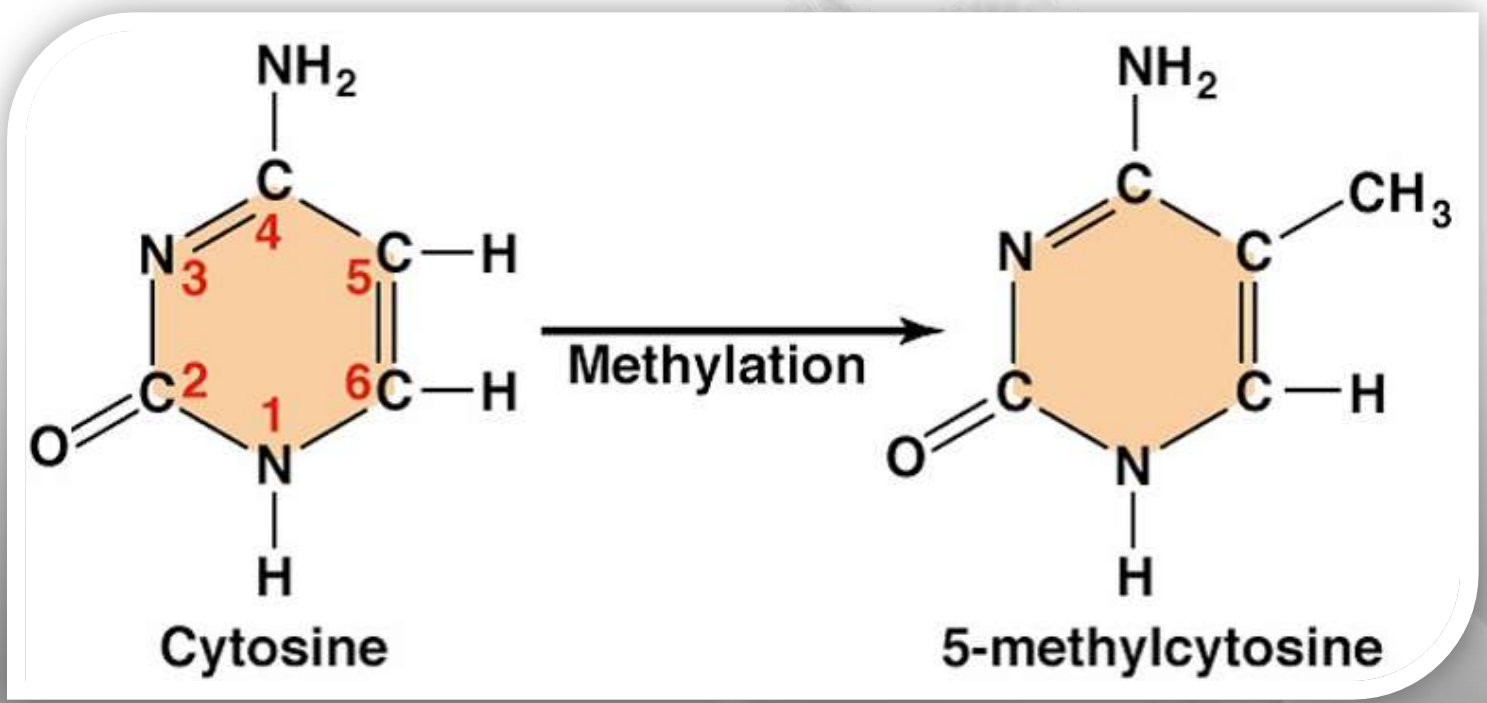
- The helix-turn-helix motif was first recognized in prokaryotic DNA-binding proteins, including the E. coli catabolite activator protein (CAP).
- In these proteins, one helix makes most of the contacts with DNA, while the other helices lie across the complex to stabilize the interaction.
- In eukaryotic cells, helix-turn-helix proteins include the homeodomain proteins, which play critical roles in the regulation of gene expression during embryonic development.
- Molecular cloning and analysis of these genes then indicated that they contain conserved sequences of 180 base pairs (called homeoboxes) that encode the DNA-binding domains (homeodomains) of transcription factors


(A) Zinc fingers
Zinc ion
b sheet
a helix
(B) Helix-turn-helix
2
1
3
(D) Helix-loop-helix
Helix
Loop
DNA-binding helix
(C) Leucine zipper
Leucine side chain
Leucine zipper domain
DNA-binding helix

- Gene Regulation at DNA Level

Chromatin Remodeling

DNA Methylation



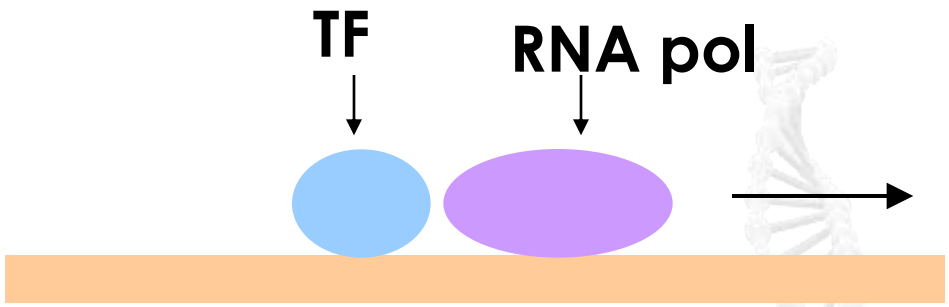
✦ **Housekeeping gene** -A gene involved in basic functions is required for the sustenance of the cell. Housekeeping genes are constitutively expressed

✦ **Luxury gene** - are those coding for specialized functions synthesized (usually) in large amounts in particular cell types.

❖ **CpG islands**

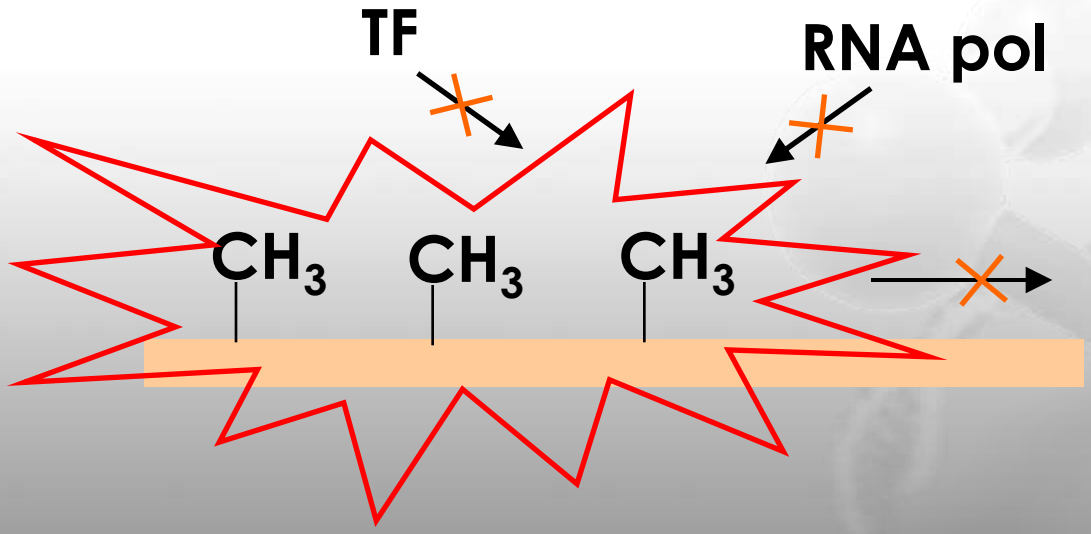
- These are genomic regions that contain a high frequency of CG dinucleotides.

- CpG islands particularly occur at or near the transcription start site of housekeeping genes.



Unmethylated CpG island

Active transcription

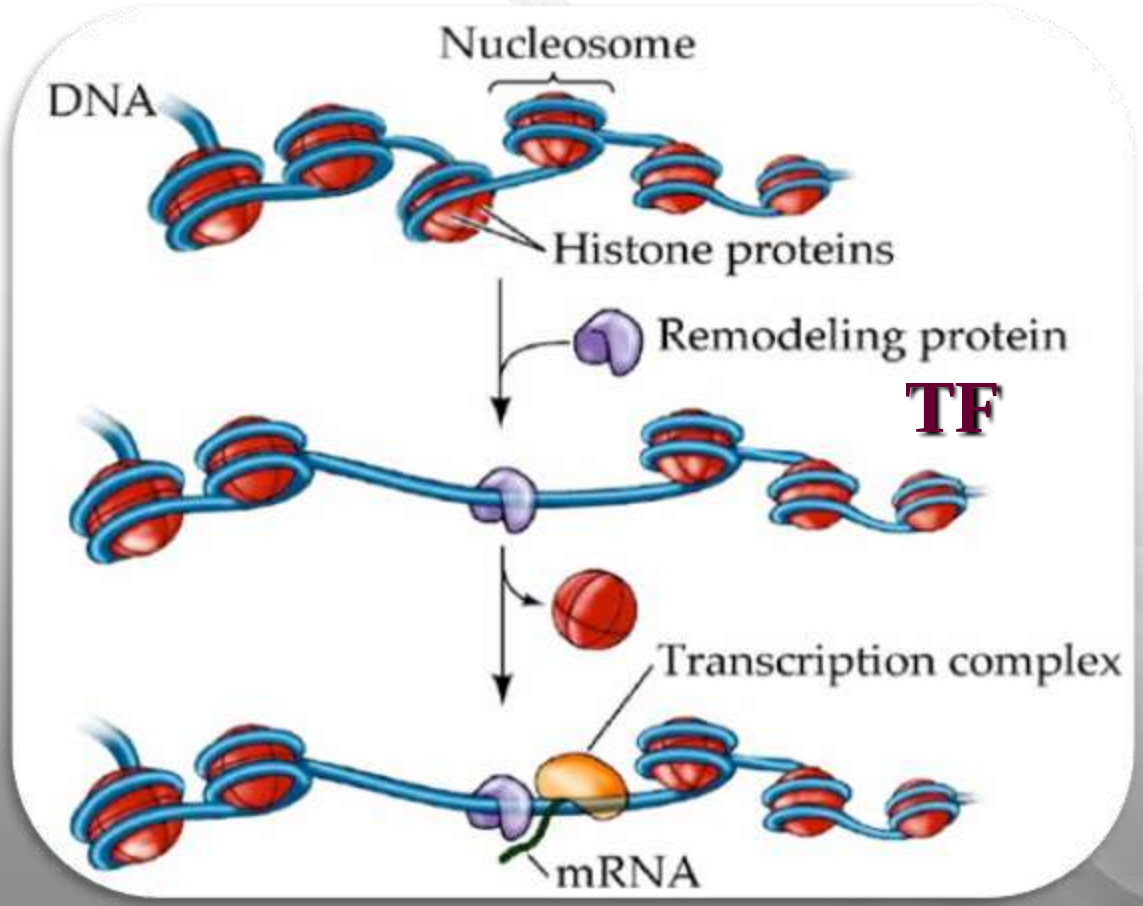


Methylated CpG island

Repressed transcription

Histone modification

- methylation
- acetylation




Thank You!