

# Regulation of Gene Expression in Eukaryotes

Md. Murad Khan  
Lecturer  
Dept. of Microbiology  
Jagannath University

# Introduction

- ▶ Gene expression is the combined process of :
  - ▶ the transcription of a gene into mRNA,
  - ▶ the processing of that mRNA, and
  - ▶ its translation into protein (for protein-encoding genes).
- ▶ **Gene expression** is the process by which the information encoded in a gene is used in the synthesis of a functional gene product.
- ▶ These products are often proteins, but in non-protein coding genes such as ribosomal RNA (rRNA), transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA.

# Introduction

- ▶ **Regulation of gene expression** includes a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (protein or RNA).
- ▶ Gene regulation is essential in both prokaryotes and eukaryotes as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed.

## Classification of genes based on expression:

- ▶ **Constitutive ( house keeping) genes:** Are expressed at a fixed rate, irrespective to the cell condition. Their structure is simpler.
- ▶ **Controllable or inducible genes:** Are expressed only as needed. Their amount may increase or decrease with respect to their basal level in different condition. Their structure is relatively complicated with some response elements

# Purposes of Regulation of Gene Expression

- ▶ Regulated expression of genes is required for
- ▶ **Adaptation:**
  - ▶ Cells of multicellular organisms respond to varying conditions. Such cells exposed to hormones and growth factors change substantially in –shape, growth rate, and other characteristics.
- ▶ **Tissue specific differentiation and development:**
  - ▶ The genetic information present in each somatic cell of a multicellular organism is practically identical.
  - ▶ Cells from muscle and nerve tissue show strikingly different morphologies and other properties, yet they contain exactly the same DNA.
  - ▶ These diverse properties are the result of differences in gene expression.
  - ▶ Expression of the genetic information is regulated during ontogeny and differentiation of the organism and its cellular components.

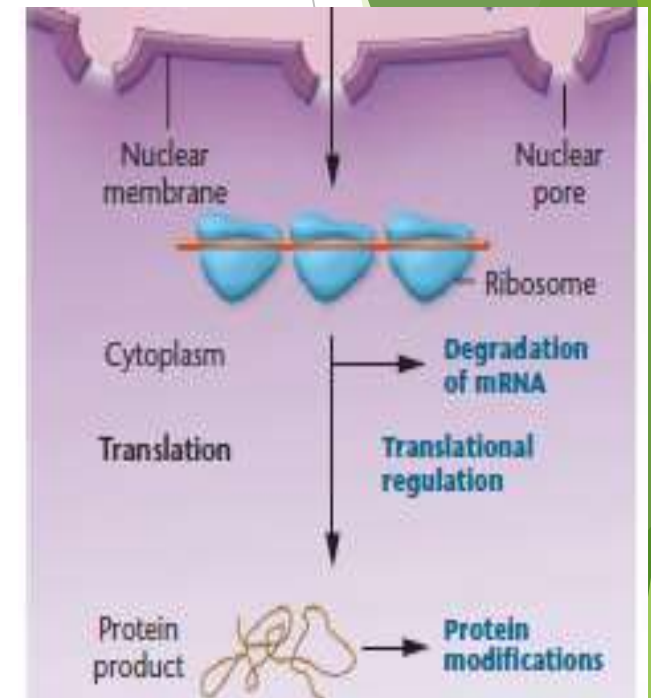
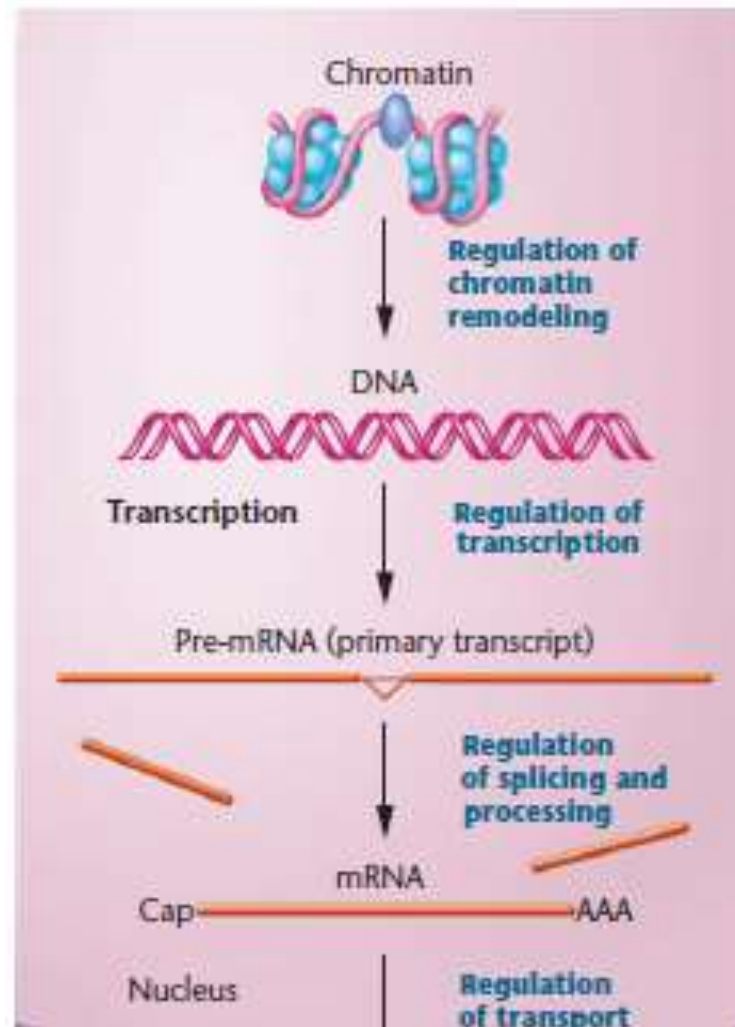
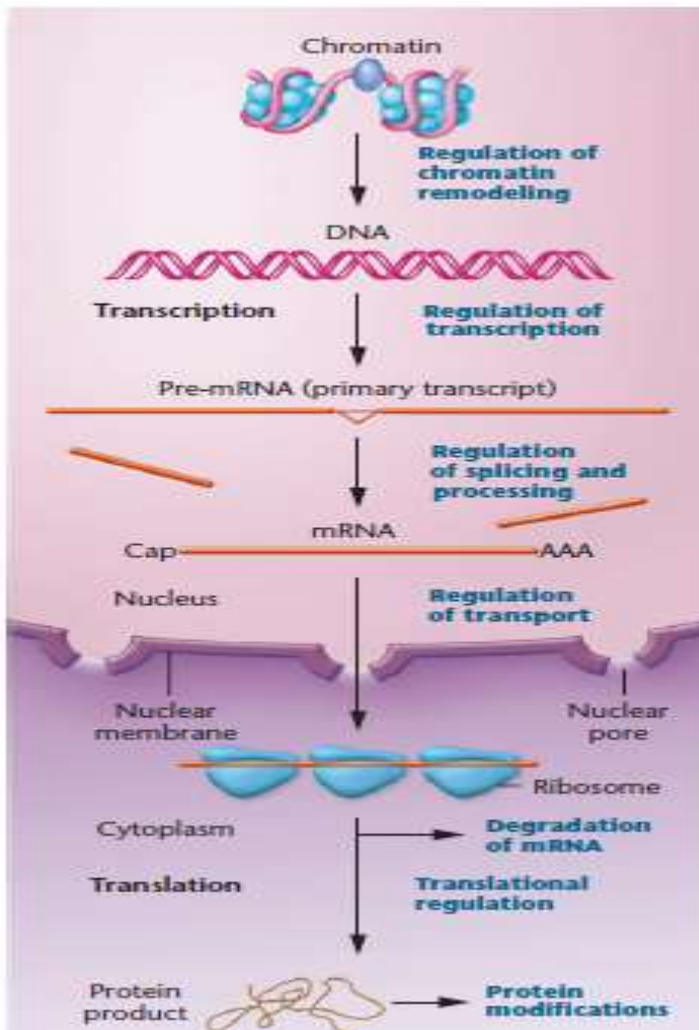
# Level of Regulation of Gene Expression

- ▶ In contrast to prokaryotic gene regulation, which occurs primarily at the level of transcription initiation, regulation of gene expression in eukaryotes can occur at many different levels. These include:
  - ▶ alterations in the DNA template,
  - ▶ initiation of transcription,
  - ▶ mRNA modifications and stability, and
  - ▶ synthesis, modification, and stability of the protein product.

# Level of Regulation of Gene Expression

- ▶ The subdivision of eukaryotic cells into organelles physically separates the events of gene expression.
- ▶ The primary event, transcription of DNA into RNA, occurs in the nucleus. RNA transcripts are also modified in the nucleus by capping, polyadenylation, and the removal of introns.
- ▶ The resulting messenger RNAs are then exported to the cytoplasm where they become associated with ribosomes, many of which are located on the membranes of the endoplasmic reticulum. Once associated with ribosomes, these mRNAs are translated into polypeptides.
- ▶ This physical separation of the events of gene expression makes it possible for regulation to occur in different places

# Level of Regulation of Gene Expression



**FIG:** Regulation can occur at any stage in the expression of genetic material in eukaryotes. All these forms of regulation affect the degree to which a gene is expressed.

# Gene Expression: Prokaryotes VS Eukaryotes

Gene regulation is significantly more complex in eukaryotes than in prokaryotes for a number of reasons:

## **Large and complex genome:**

- ▶ Eukaryotic cells contain a much greater amount of DNA than do prokaryotic cells, and this DNA is associated with histones and other proteins to form highly compact chromatin structures within an enclosed nucleus. Eukaryotic cells modify this structural organization in order to influence gene expression.
- ▶ The *E. coli* genome consists of a single, circular chromosome containing 4.6 Mb. This genome encodes approximately 2000 proteins.



# Gene Expression: Prokaryotes VS Eukaryotes

- ▶ In comparison, the genome within a human cell contains 23 pairs of chromosomes ranging in size from 50 to 250 Mb. Approximately 40,000 genes are present within the 3000 Mb of human DNA. It would be very difficult for a DNA-binding protein to recognize a unique site in this vast array of DNA sequences. More-elaborate mechanisms are required to achieve specificity.
- ▶ Eukaryotic genes are located on many chromosomes (rather than just one), and these chromosomes are enclosed within a double-membrane-bound nucleus. After transcription, transport of RNAs into the cytoplasm can be regulated in order to modulate the availability of mRNAs for translation.

# Gene Expression: Prokaryotes VS Eukaryotes

## **Uncoupled transcription and translation:**

- ▶ In prokaryotes, transcription and translation are coupled processes, the primary transcript is immediately translated.
- ▶ The transcription and translation are uncoupled in eukaryotes, eliminating some potential gene-regulatory mechanisms. The primary transcript in eukaryotes undergoes modifications to become a mature functional m RNA.

# Gene Expression: Prokaryotes VS Eukaryotes

## **mRNA splicing, modification and half-life:**

- ▶ The mRNAs of most eukaryotic genes must be spliced, capped, and poly-adenylated prior to transport from the nucleus. Each of these processes can be regulated in order to influence the numbers and types of mRNAs available for translation.
- ▶ Eukaryotic mRNAs can have a wide range of half-lives. In contrast, the majority of prokaryotic mRNAs decay very rapidly. Rapid turnover of mRNAs allows prokaryotic cells to rapidly respond to environmental changes. In eukaryotes, the types and quantities of mRNAs in each cell type can be more subtly manipulated by altering mRNA decay rates over a larger range.

# Gene Expression: Prokaryotes VS Eukaryotes

## **Different cell types:**

- ▶ Different cell types are present in most eukaryotes. Liver and pancreatic cells, for example, differ dramatically in the genes that are highly expressed. Different mechanisms are involved in the regulation of such genes.

## **Absence of operons:**

- ▶ The eukaryotic genes are not generally organized into operons as are there in prokaryotes. Instead, genes that encode proteins for steps within a given pathway are often spread widely across the genome.

## **Chromatin structure:**

- ▶ The DNA in eukaryotic cells is extensively folded and packed into the protein-DNA complex called chromatin. Histones are an important part of this complex since they both form the structures known as nucleosomes and also contribute significantly into gene regulatory mechanisms.

# Spatiotemporal Gene Expression

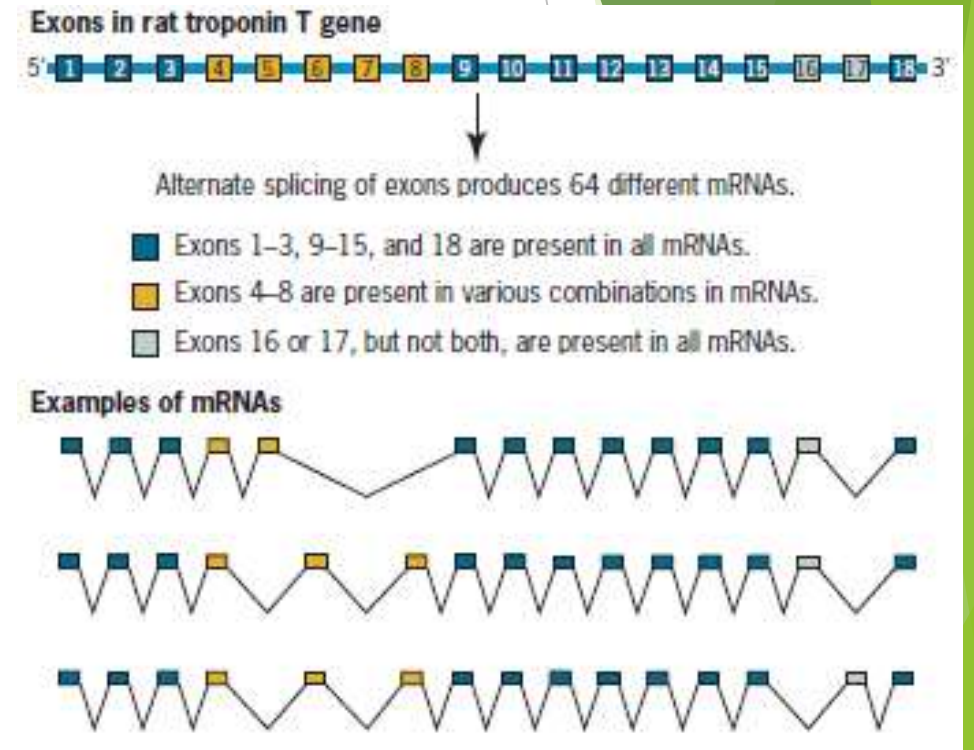
- ▶ Spatiotemporal gene expression is the activation of genes within specific tissues of an organism at specific times during development. Spatiotemporal variation plays a key role in generating the diversity of cell types found in developed organisms; since the identity of a cell is specified by the collection of genes actively expressed within that cell, if gene expression was uniform spatially and temporally, there could be at most one kind of cell.
- ▶ All cells have the same genes, but not all genes are expressed in every cell at all times. For example, muscle cells contain the gene for the pancreatic hormone insulin; yet muscles do not produce insulin. Similarly, genes that produce certain hormones or nerve signals may only be expressed in the presence of specific stimulus from the environment.
- ▶ **Spatial gene expression**– not every gene product needed in every cell type.
- ▶ **Temporal gene expression** – different genes expressed at different times.

# Alternate RNA Splicing

- ▶ Most eukaryotic genes possess introns, noncoding regions that interrupt the sequence that specifies the amino acids of a polypeptide. Each intron must be removed from the RNA transcript of a gene in order for the coding sequence to be expressed properly.
- ▶ Genes with multiple introns present a curious problem to the RNA splicing machinery. These introns can be removed separately or in combination, depending on how the splicing machinery interacts with the RNA.
- ▶ If two successive introns are removed together, the exon between them will also be removed. Thus, the splicing machinery has the opportunity to modify the coding sequence of an RNA by deleting some of its exons.
- ▶ This phenomenon of splicing an RNA transcript in different ways is apparently a way of economizing on genetic information. Instead of duplicating genes, or pieces of genes, the alternate splicing of transcripts makes it possible for a single gene to encode different polypeptides.

# Alternate RNA Splicing

- ▶ One example of alternate splicing occurs during the expression of the gene for troponin T, a protein found in the skeletal muscles of vertebrates; the size of this protein ranges from about 150 to 250 amino acids. In the rat, the troponin T gene is more than 16 kb long and contains 18 different exons. Transcripts of this gene are spliced in different ways to create a large array of mRNAs. When these are translated, many different troponin T polypeptides are produced.
- ▶ All these polypeptides share amino acids from exons 1–3, 9–15, and 18. However, the regions encoded by exons 4–8 may be present or absent, depending on the splicing pattern, and apparently in any combination. Additional variation is provided by the presence or absence of regions encoded by exons 16 and 17; if 16 is present, 17 is not, and vice versa.



**FIG:** Alternate splicing of transcripts from the rat troponin T gene. Only 3 of the possible 64 different mRNAs are shown.



# Cytoplasmic Control of mRNA Stability

- ▶ An mRNA that is rapidly degraded must be replenished by additional transcription; otherwise, the polypeptide it encodes will cease to be synthesized. This cessation of polypeptide synthesis may, of course, be part of a developmental program. Once the polypeptide has had its effect, it may no longer be needed; in fact, its continued synthesis may be harmful.

## **Messenger RNA longevity can be influenced by several factors:**

- ▶ Poly(A) tails seem to stabilize mRNAs. If the poly-A tail is shortened to less than about 30 nucleotides, the mRNA becomes unstable and acts as a substrate for exonucleases that degrade the RNA in either a 5' to 3' or 3' to 5' direction.
- ▶ The sequence of the 3' untranslated region (3' UTR) preceding a poly(A) tail also seems to affect mRNA stability. Several short-lived mRNAs have the sequence AUUUA repeated several times in their 3' untranslated regions. When this sequence is artificially transferred to the 3' untranslated region of more stable mRNAs, they, too, become unstable.



# Cytoplasmic Control of mRNA Stability

- ▶ Decapping enzymes can remove the 7-methylguanosine cap, which also renders the mRNA unstable.
- ▶ An mRNA may be cleaved internally by an endonuclease, providing unprotected ends at which exonuclease degradation may proceed.
- ▶ Chemical factors, such as hormones, may also affect mRNA stability.
- ▶ Stability of mRNAs and the translation of mRNAs into polypeptides are also regulated by small, noncoding RNA molecules called small interfering RNAs (siRNAs) or microRNAs (miRNAs). Short interfering and microRNAs base-pair with sequences in specific mRNAs; once paired, they either cause the mRNA to be cleaved and subsequently degraded, or they prevent the mRNA from being translated into a polypeptide.

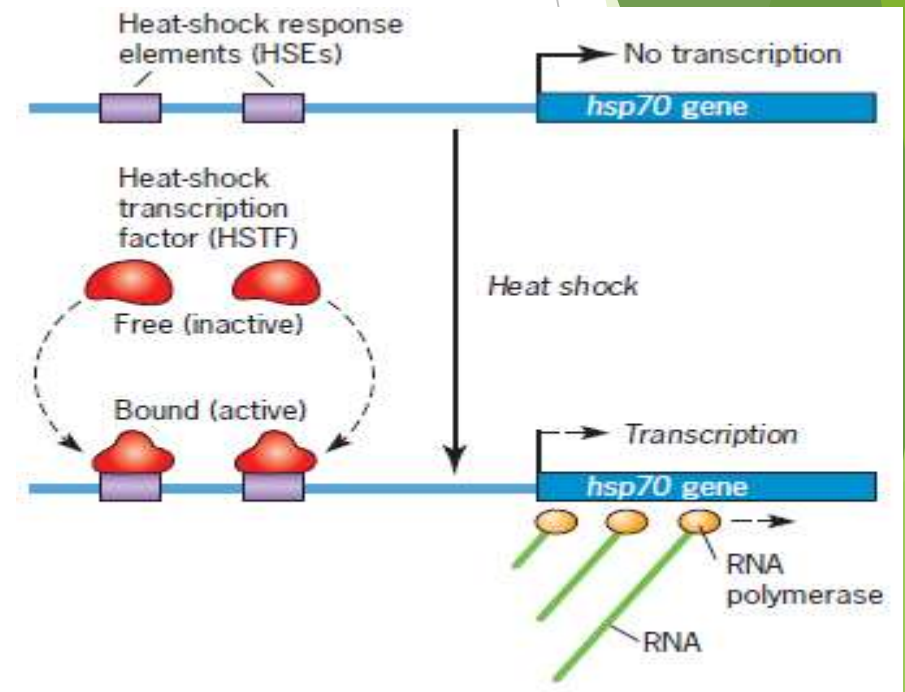
# Induction of Transcriptional Activity by Environmental Factors

## Temperature: The heat shock gene

- ▶ When organisms are subjected to the stress of high temperature, they respond by synthesizing a group of proteins, called **heat-shock proteins**, that help to stabilize the internal cellular environment. The expression of the heat-shock proteins is regulated at the transcriptional level; transcription of the heat-shock genes is induced by heat.
- ▶ In *Drosophila*, for example, one of the heat-shock proteins called HSP70 (for heat-shock protein, molecular weight 70 kilodaltons) is encoded by a family of genes located in two nearby clusters on one of the autosomes. Altogether, there are five to six copies of these hsp70 genes in the two clusters. When the temperature exceeds 33°C, as it does on hot summer days, each of the genes is transcribed into RNA, which is then processed and translated to produce HSP70 polypeptides.

# Induction of Transcriptional Activity by Environmental Factors

- ▶ This heat induced transcription of the *hsp70* genes is mediated by a polypeptide called the heat-shock transcription factor, or HSTF, which is present in the nuclei of *Drosophila* cells. When *Drosophila* are heat stressed, the HSTF is chemically altered by phosphorylation. In this altered state, it binds specifically to nucleotide sequences upstream of the *hsp70* genes and makes the genes more accessible to RNA polymerase II, the enzyme that transcribes most protein-encoding genes. The transcription of the *hsp70* genes is then vigorously stimulated.
- ▶ The sequences to which the phosphorylated HSTF binds are called heat-shock response elements (HSEs).



**FIG:** Induction of transcription from the *Drosophila hsp70* gene by heat shock.

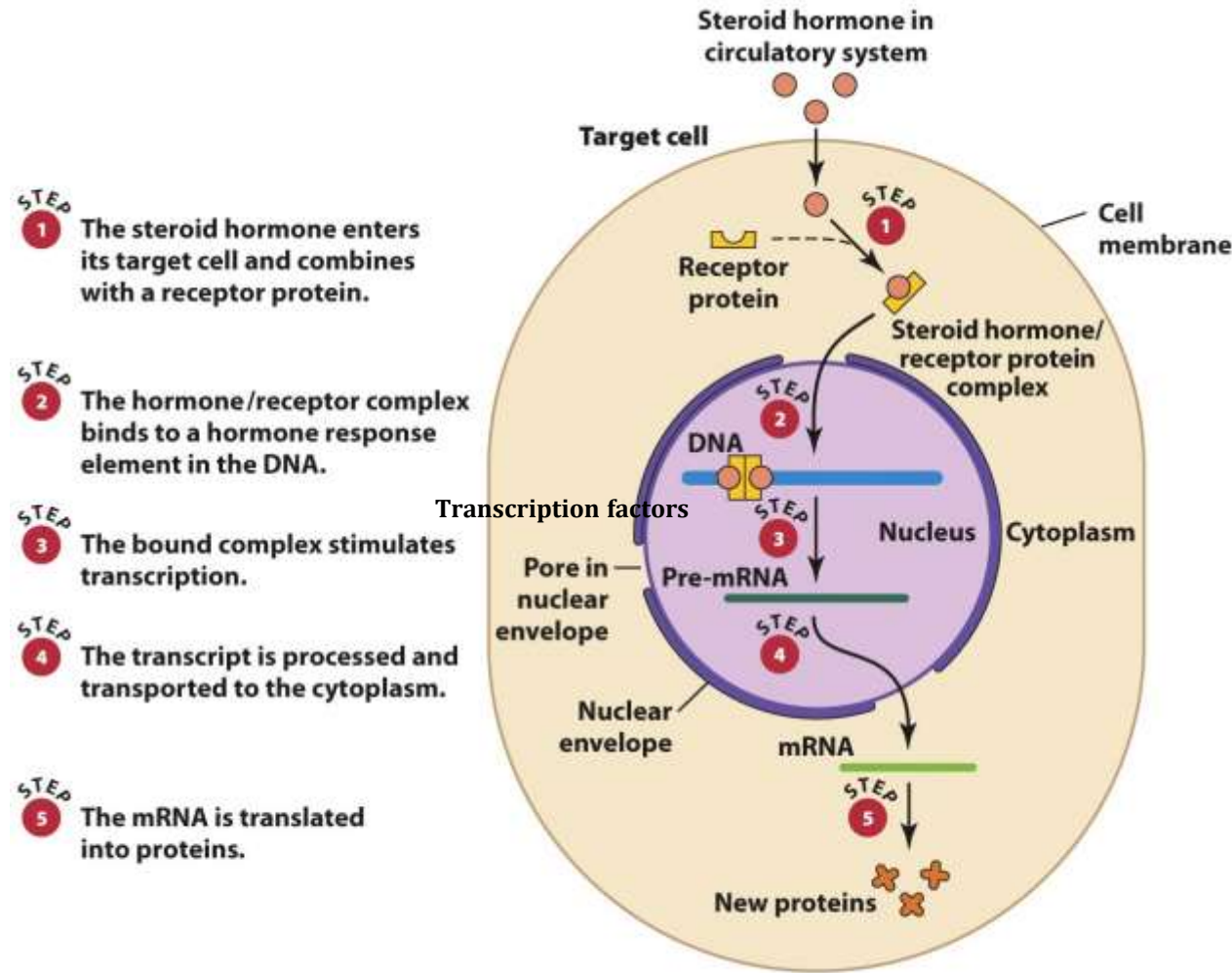
# Induction of Transcriptional Activity by Biological Factors

## **Signal molecules: Genes that response by hormones**

- ▶ In multicellular eukaryotes, one type of cell can signal another by secreting a hormone. Hormones circulate through the body, make contact with their target cells, and then initiate a series of events that regulate the expression of particular genes.

## Regulation of gene expression by steroid hormones

- ▶ Steroid hormones are small, lipid-soluble molecules derived from cholesterol. Because of their lipid nature, they have little or no trouble passing through cell membranes.
- ▶ Once these hormones have entered a cell, they interact with cytoplasmic or nuclear proteins called hormone receptors. The receptor/hormone complex that is formed then interacts with the DNA where it acts as a transcription factor to regulate the expression of certain genes.



© 2012 John Wiley & Sons, Inc. All rights reserved.

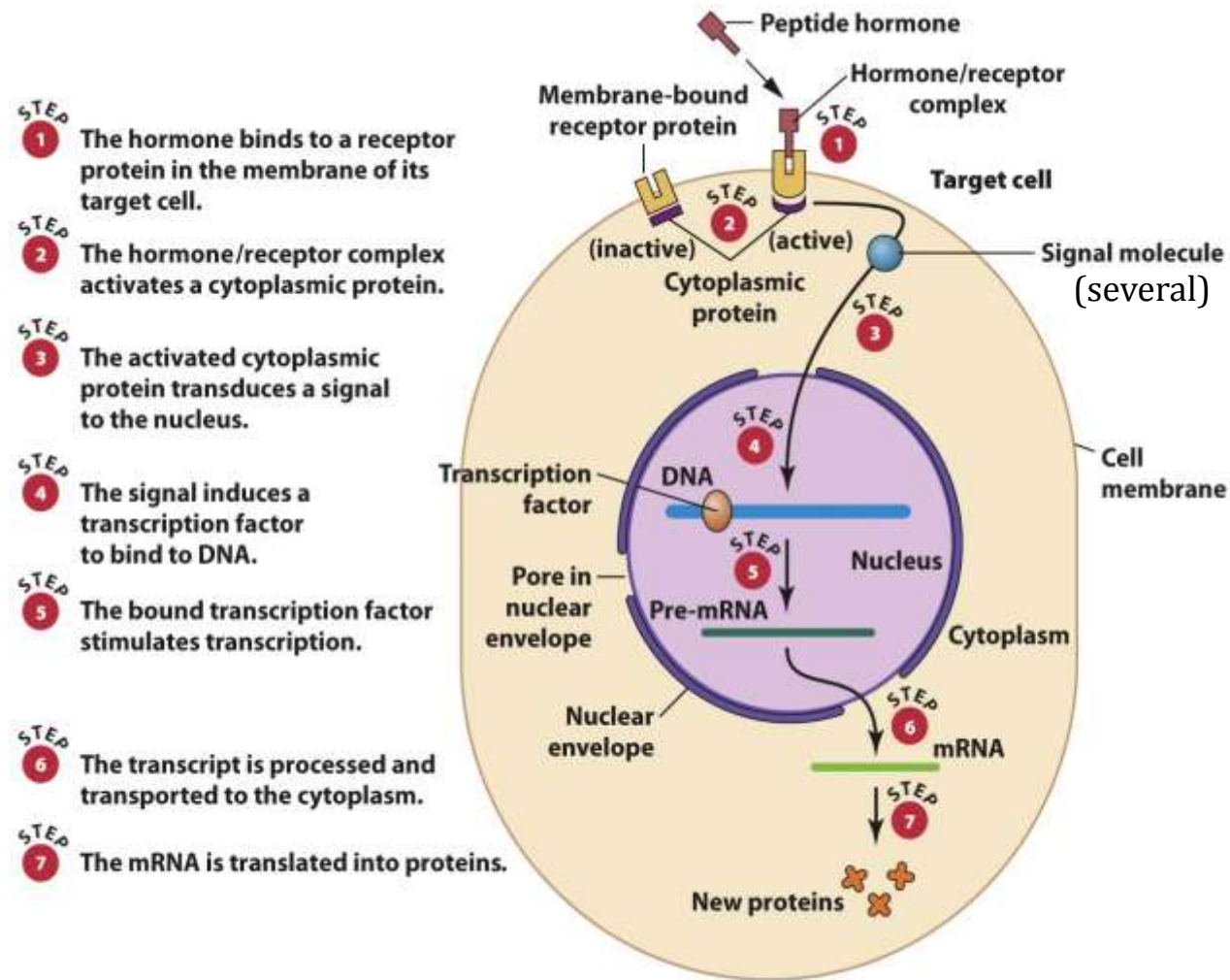
**FIG:** Regulation of gene expression by steroid hormones. The hormone interacts with a receptor inside its target cell. In this example the receptor is in the cytoplasm; other steroid hormone receptors are located in the nucleus. The steroid/ hormone receptor complex moves into the nucleus where it activates the transcription of particular genes.

# Induction of Transcriptional Activity by Biological Factors

## Regulation of gene expression by peptide hormones

- ▶ Peptide hormones, are linear chains of amino acids. Like all other polypeptides, these molecules are encoded by genes. Because peptide hormones are typically too large to pass freely through cell membranes, the signals they convey must be transmitted to the interior of cells by membrane-bound receptor proteins.
- ▶ When a peptide hormone interacts with its receptor, it causes a conformational change in the receptor that eventually leads to changes in other proteins inside the cell. Through a cascade of such changes, the hormonal signal is transmitted through the cytoplasm of the cell and into the nucleus, where it ultimately has the effect of regulating the expression of specific genes. This process of transmitting the hormonal signal through the cell and into the nucleus is called **signal transduction**.





© 2012 John Wiley & Sons, Inc. All rights reserved.

**FIG:** Regulation of gene expression by peptide hormones. The hormone (an extracellular signal) interacts with a receptor in the membrane of its target cell. The resulting hormone/receptor complex activates a cytoplasmic protein that triggers a cascade of intracellular changes. These changes transmit the signal into the nucleus, where a transcription factor stimulates the expression of particular genes.

# Induction of Transcriptional Activity by Biological Factors

- ▶ Hormone-induced gene expression is mediated by specific sequences in the DNA. These sequences are called **hormone response elements (HREs)**.
- ▶ They are situated near the genes they regulate and serve to bind specific proteins, which then act as transcription factors. With steroid hormones such as estrogen, the HREs are bound by the hormone/ receptor complex, which then stimulates transcription. The vigor of this transcriptional response depends on the number of HREs present. When there are multiple response elements, hormone/receptor complexes bind cooperatively with each other, significantly increasing the rate of transcription; that is, a gene with two response elements is transcribed more than twice as vigorously as a gene with only one.
- ▶ With peptide hormones, the receptor usually remains in the cell membrane, even after it has formed a complex with the hormone. The hormonal signal is therefore conveyed to the nucleus by other proteins, some of which bind to sequences near the genes that are regulated by the hormone. These proteins then act as transcription factors to control the expression of the genes.



# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ The transcription of eukaryotic genes is regulated by interactions between proteins and DNA sequences within or near the genes.
- ▶ **Cis-acting elements** - DNA sequences close to a gene that are required for gene expression.
  - ▶ Enhancers
  - ▶ Promoters
- ▶ **Trans acting elements/factors** - factors, usually considered to be proteins, that bind to the cis-acting sequences to control gene expression.
  - ▶ RNA polymerase
  - ▶ Transcription factors
  - ▶ Domains of trans-acting factors

# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ **Enhancer** elements are DNA sequences, although they have no promoter activity of their own but they greatly increase the activities of many promoters in eukaryotes.
- ▶ Enhancers function by serving as binding sites for specific regulatory proteins.
- ▶ An enhancer is effective only in the specific cell types in which appropriate regulatory proteins are expressed.

## Properties of Enhancer:

- ▶ Enhancer elements can exert their positive influence on transcription even when separated by thousands of base pairs from a promoter.
- ▶ They work when oriented in either direction; and they can work upstream (5') or downstream (3') from the promoter.
- ▶ The effects of enhancers are independent of position. They may be upstream, downstream, or within an intron.
- ▶ Enhancers are promiscuous; they can stimulate any promoter in the vicinity and may act on more than one promoter.

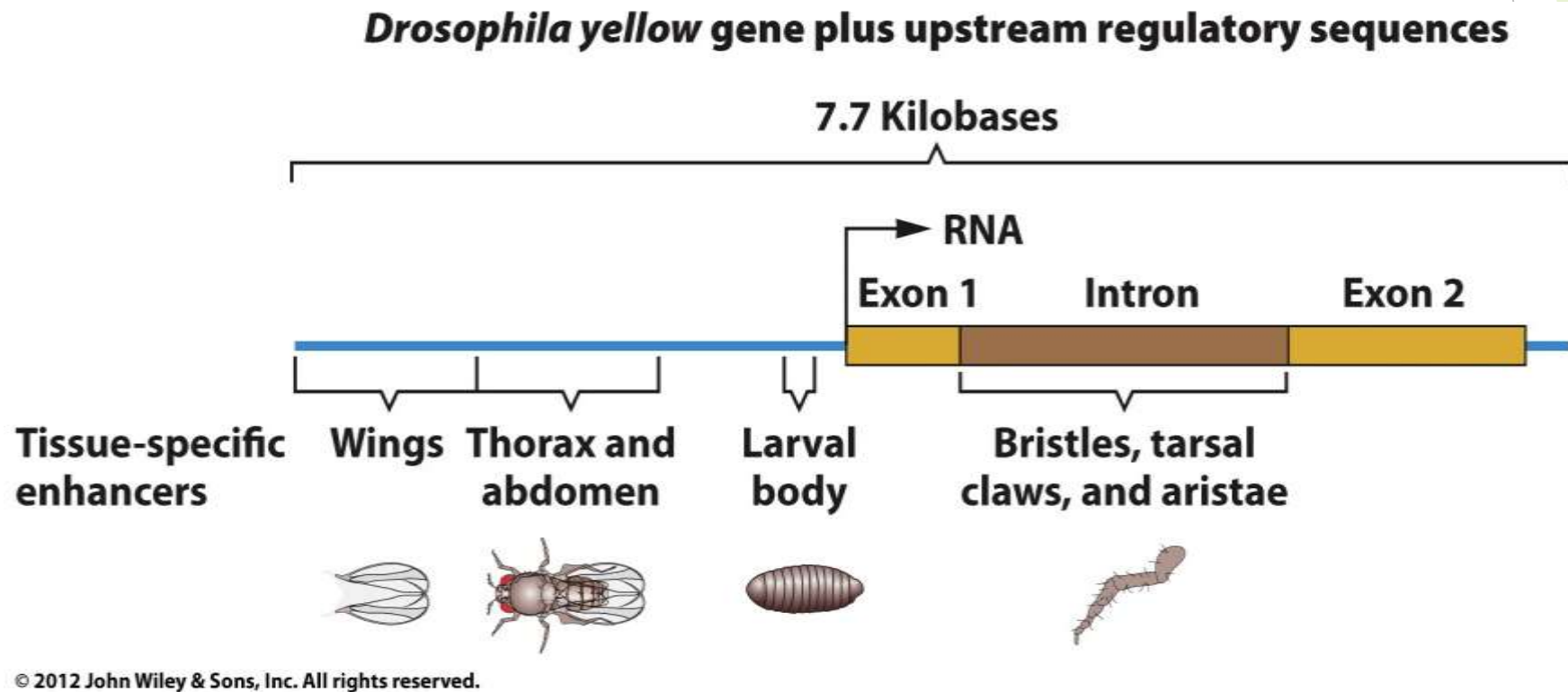
# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ The elements that decrease or repress the expression of specific genes have also been identified.
- ▶ **Silencers** are control regions of DNA that, like enhancers, may be located thousands of base pairs away from the gene they control.
- ▶ However, when transcription factors bind to them, expression of the gene they control is repressed.

# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ Tissue-specific gene expression is mediated by enhancers or enhancer-like elements. A clear example of this tissue specificity comes from the study of the yellow gene in *Drosophila*. This gene is responsible for pigmentation in many parts of the body—in the wings, legs, thorax, and abdomen.
- ▶ Wild-type flies show a dark brownish-black pigment in all these structures, whereas mutant flies show a lighter yellowish-brown pigment. However, in some mutants, there is a mosaic pattern of pigmentation, brownish-black in some tissues and yellowish-brown in others. These mosaic patterns are due to mutations that alter the transcription of the yellow gene in some tissues but not in others.
- ▶ If, for example, the enhancer for expression in the wing is mutated, the bristles on the wings are yellowish-brown instead of brownish-black. The battery of enhancers associated with the yellow gene allows its expression to be controlled in a tissue-specific way.

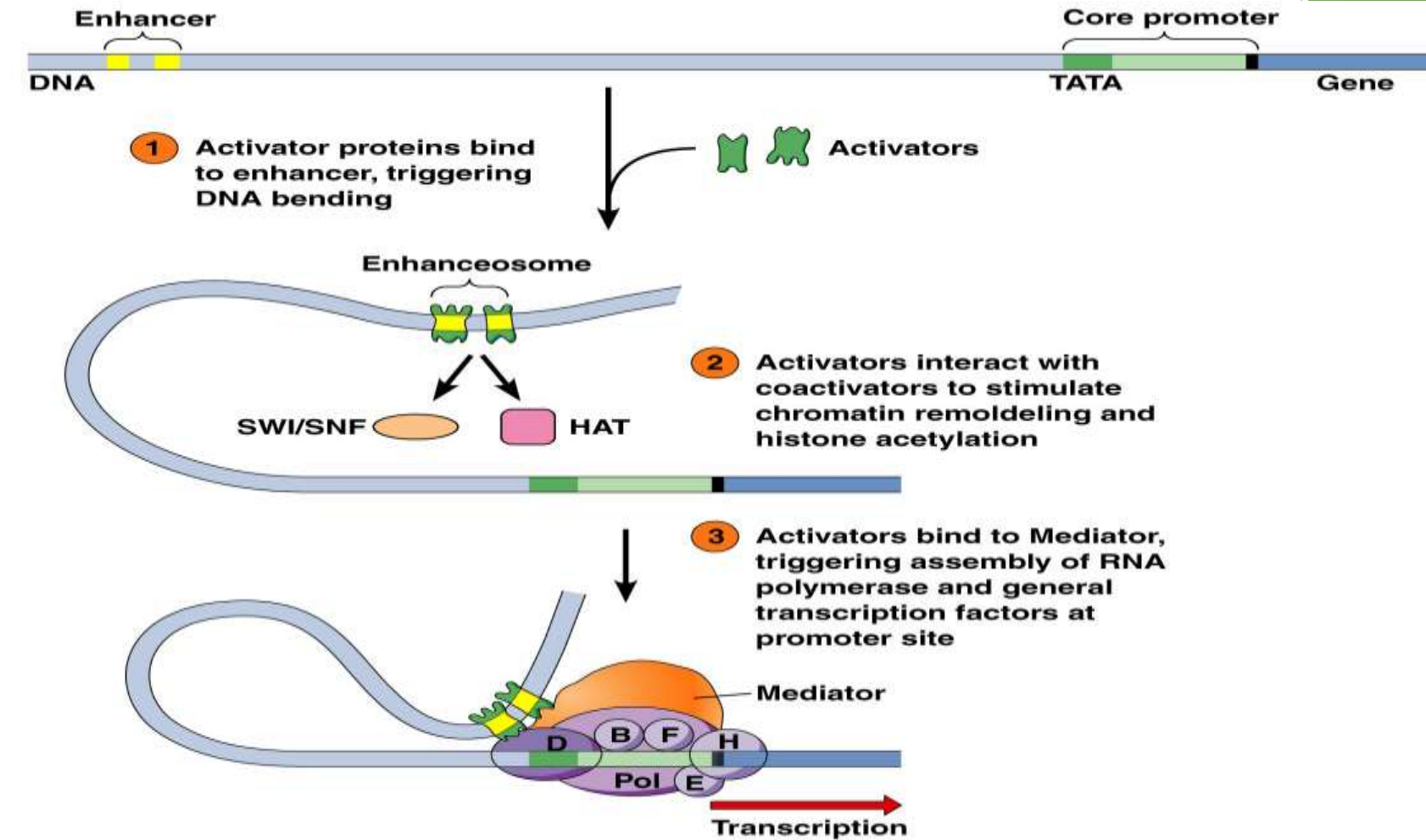
# DNA Sequences and Proteins Involved in the Control of Transcription



**FIG:** The tissue-specific enhancers of the *Drosophila yellow* gene.

# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ **Basal transcription factors** are proteins that bind to specific DNA sequences within the promoter to facilitate RNA polymerase alignment.
- ▶ **Special transcription factors** are proteins that bind to response elements or to sequences called enhancers that are located near a gene and facilitate the action of basal transcription factors and RNA polymerase.
- ▶ Proteins that bind to **enhancers** influence the activity of proteins that bind to promoters, including the **basal transcription factors** and **RNA polymerase**.
- ▶ These proteins are brought into contact with one another by the **mediator complex** (at least 20 different proteins). In this way, then, proteins bound to the enhancer exert control over transcription, which is initiated at the promoter.

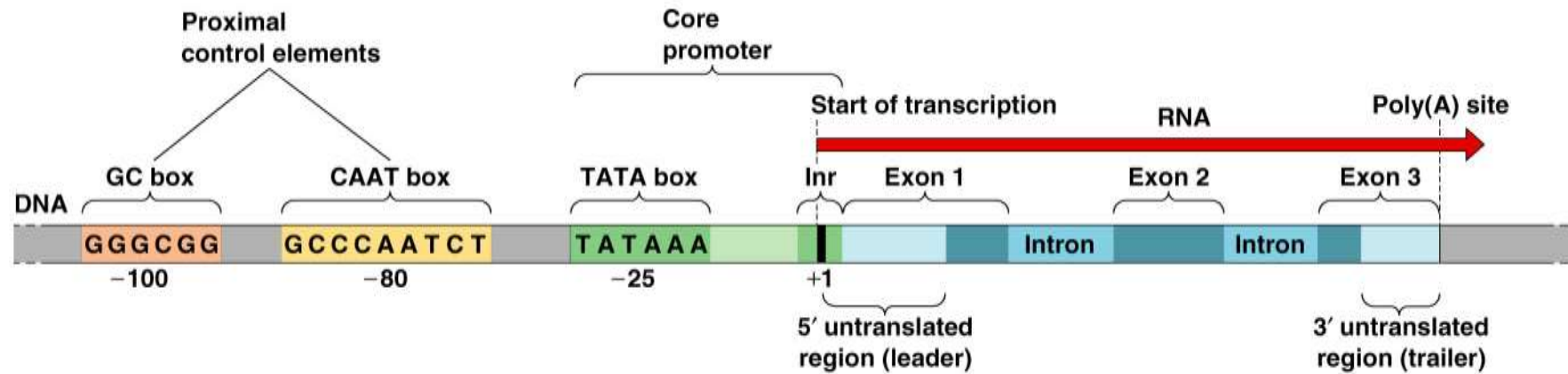


**FIG:** Model for enhancer action.

# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ **Promoter** consists of short nucleotide sequence that serve as the recognition point for binding of RNA polymerase.
- ▶ Located immediately adjacent to the genes they regulate, upstream from the transcription start point.
- ▶ Promoters for RNA polymerase II include:
  - ▶ TATA box
  - ▶ CAAT box
  - ▶ GC box
  - ▶ Octamer box
- ▶ Promoters for RNA polymerase I & III have a different sequence and bind different transcription factors.





**FIG:** Anatomy of a typical eukaryotic gene, with its core promoter and proximal control region.

|                    | Site  | Structure  | Importance   |
|--------------------|---|--|--|
| <b>TATA box</b>    | 25-30 bp upstream (from the initial point of transcription)   | 8 bp sequences composed only of T=A pairs.         | Mutations in this sequence greatly reduce transcription (Loosing the ability to bind to transcription factors) |
| <b>CAAT box</b>    | 70-80 bp upstream (from the initial point of transcription)   | CAAT or CCAAT sequence.                            | Mutations in this sequence greatly reduce transcription  |
| <b>GC box</b>      | 110 bp upstream (from the initial point of transcription)     | GGGCGG sequence, often present in multiple copies. | Documented by mutational analysis  |
| <b>Octamer box</b> | 120-130 bp upstream (from the initial point of transcription) | ATTTGCAT sequence.                                 | Affects the efficiency of promoter in initiating transcription.  |

# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ Transcription factors usually have two domains (fragment) that may be in separate parts of the molecule or overlapping.
  - ▶ A DNA binding domain
  - ▶ A transcriptional activation domain
- ▶ Transcriptional activation appears to involve physical interactions between proteins. A transcription factor that has bound to an enhancer may make contact with one or more proteins at other enhancers, or it may interact directly with proteins that have bound in the promoter region. Through these contacts and interactions, the transcriptional activation domain of the factor may then induce conformational changes in the assembled proteins, paving the way for the RNA polymerase to initiate transcription.

# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ Many eukaryotic transcription factors have characteristic structural motifs that result from associations between amino acids within their polypeptide chains.
- ▶ **Zinc finger:** a short peptide loop that forms when two cysteines in one part of the polypeptide and two histidines in another part nearby jointly bind a zinc ion; the peptide segment between the two pairs of amino acids then juts out from the main body of the protein as a kind of finger.
- ▶ **Helix-turn-helix:** a stretch of three short helices of amino acids separated from each other by turn. Helical segment closest to the carboxy terminus is required for DNA binding; the other helices seem to be involved in the formation of protein dimers.

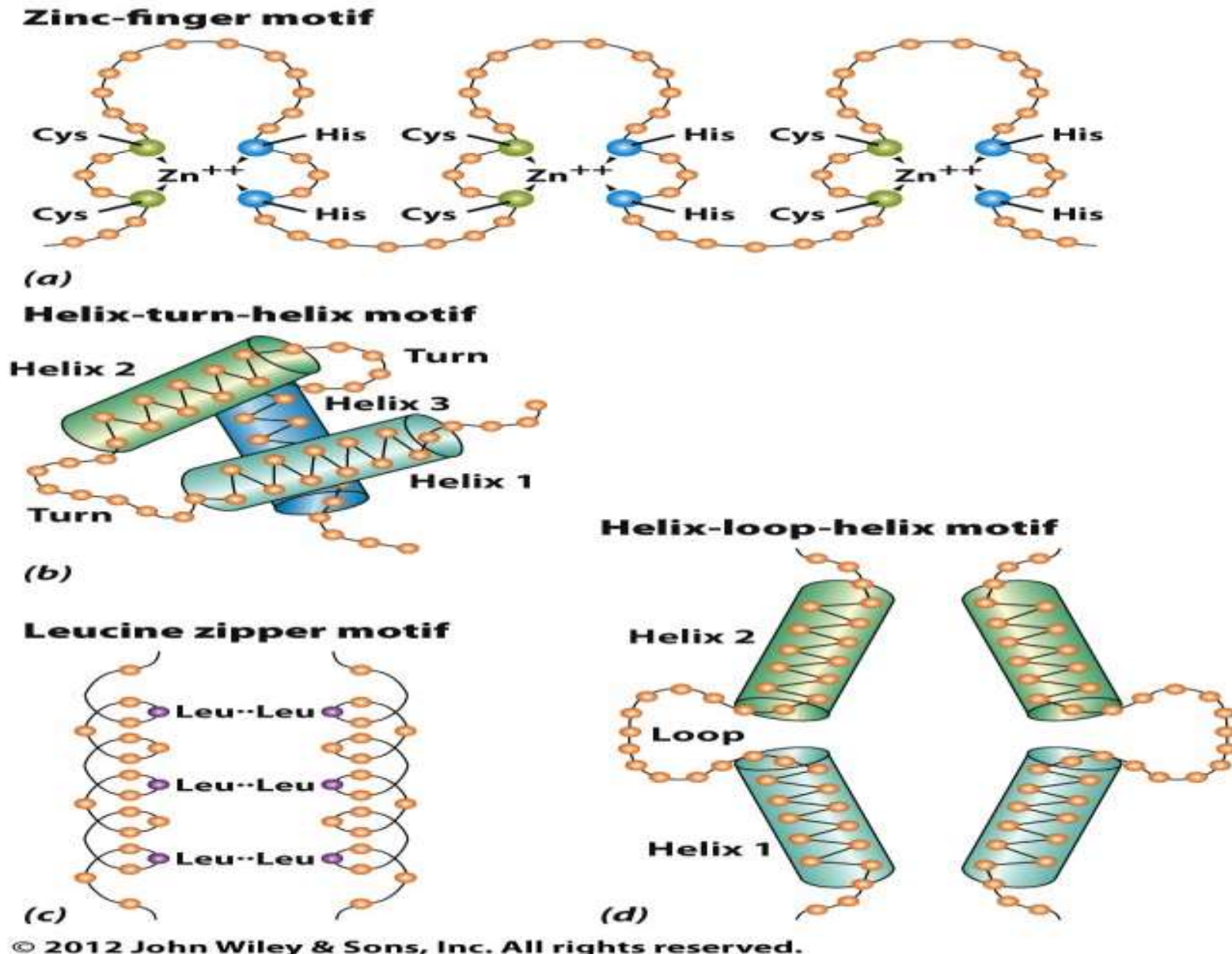
# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ **Leucine zipper:** a stretch of amino acids with a leucine at every seventh position. Polypeptides with this feature can form dimers by interactions between the leucines in each of their zipper regions. Usually, the zipper sequence is adjacent to a positively charged stretch of amino acids. When two zippers interact, these charged regions splay out in opposite directions, forming a surface that can bind to negatively charged DNA.
- ▶ **Helix-loop-helix:** a stretch of two helical regions of amino acids separated by a nonhelical loop. The helical regions permit dimerization between two polypeptides. Sometimes the helix-loop-helix motif is adjacent to a stretch of basic (positively charged) amino acids, so that when dimerization occurs, these amino acids can bind to negatively charged DNA.



# DNA Sequences and Proteins Involved in the Control of Transcription

- ▶ Transcription factors with dimerization motifs such as the leucine zipper or the helix-loop-helix could, in principle, combine with polypeptides like themselves to form homodimers, or they could combine with different polypeptides to form heterodimers.



**FIG:** Structural motifs within different types of transcription factors. (a) Zinc-finger motifs in the mammalian transcription factor SP1. (b) Helix-turn-helix motif in a homeodomain transcription factor. (c) A leucine zipper motif that allows two polypeptides to dimerize and then bind to DNA. (d) A helix-loop-helix motif that allows two polypeptides to dimerize and then bind to DNA.

# Regulation of Gene Expression by RNA Interference

- ▶ **Short noncoding RNAs** may regulate the expression of eukaryotic genes by interacting with the messenger RNAs produced by these genes.
- ▶ By base-pairing with target sequences in messenger RNA molecules, these small RNAs interfere with gene expression. Hence, this type of posttranscriptional gene regulation is called **RNA interference**, often abbreviated as RNAi.
- ▶ The phenomenon of RNA interference involves small RNA molecules called **short interfering RNAs (siRNAs)** or **microRNAs (miRNAs)**. These molecules, 21 to 28 base pairs long, are produced from larger, double-stranded RNA molecules by the enzymatic action of proteins that are double stranded RNA-specific endonucleases. Because these endonucleases “dice” large RNA into small pieces, they are called **Dicer enzymes**.
- ▶ The siRNAs and miRNAs produced by Dicer activity are base-paired throughout their lengths except at their 3' ends, where two nucleotides are unpaired.

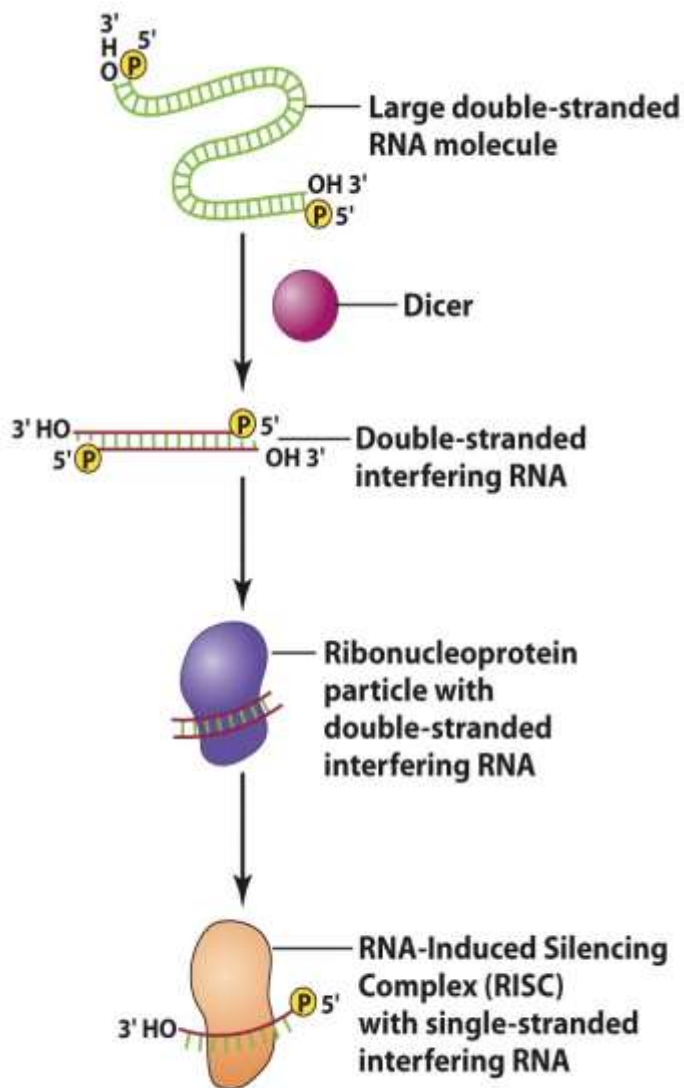
# Regulation of Gene Expression by RNA Interference

- ▶ In the cytoplasm, siRNAs and miRNAs become incorporated into ribonucleoprotein particles. The double-stranded siRNA or miRNA in these particles is unwound, and one of its strands is preferentially eliminated.
- ▶ The surviving single strand of RNA is then able to interact with specific messenger RNA molecules. This interaction is mediated by base-pairing between the single strand of RNA in the RNA–protein complex and a complementary sequence in the messenger RNA molecule.
- ▶ Because this interaction prevents the expression of the gene that produced the mRNA, the RNA–protein particle is called an **RNA-Induced Silencing Complex (RISC)**.

# Regulation of Gene Expression by RNA Interference

- ▶ Whenever the base-pairing between the RNA within the RISC and the target sequence in the mRNA is perfect or nearly so, the RISC cleaves the target mRNA in the middle of the base paired region. The cleaved mRNA is then degraded. After cleavage, the RISC may associate with another molecule of mRNA and induce its cleavage. RISC-associated RNAs that result in mRNA cleavage are usually termed **short interfering RNAs**.
- ▶ Whenever the RNA within the RISC pairs imperfectly with its target sequence, the mRNA is usually not cleaved; instead, translation of the mRNA is inhibited. RISC-associated RNAs that have this effect are usually termed **microRNAs**.

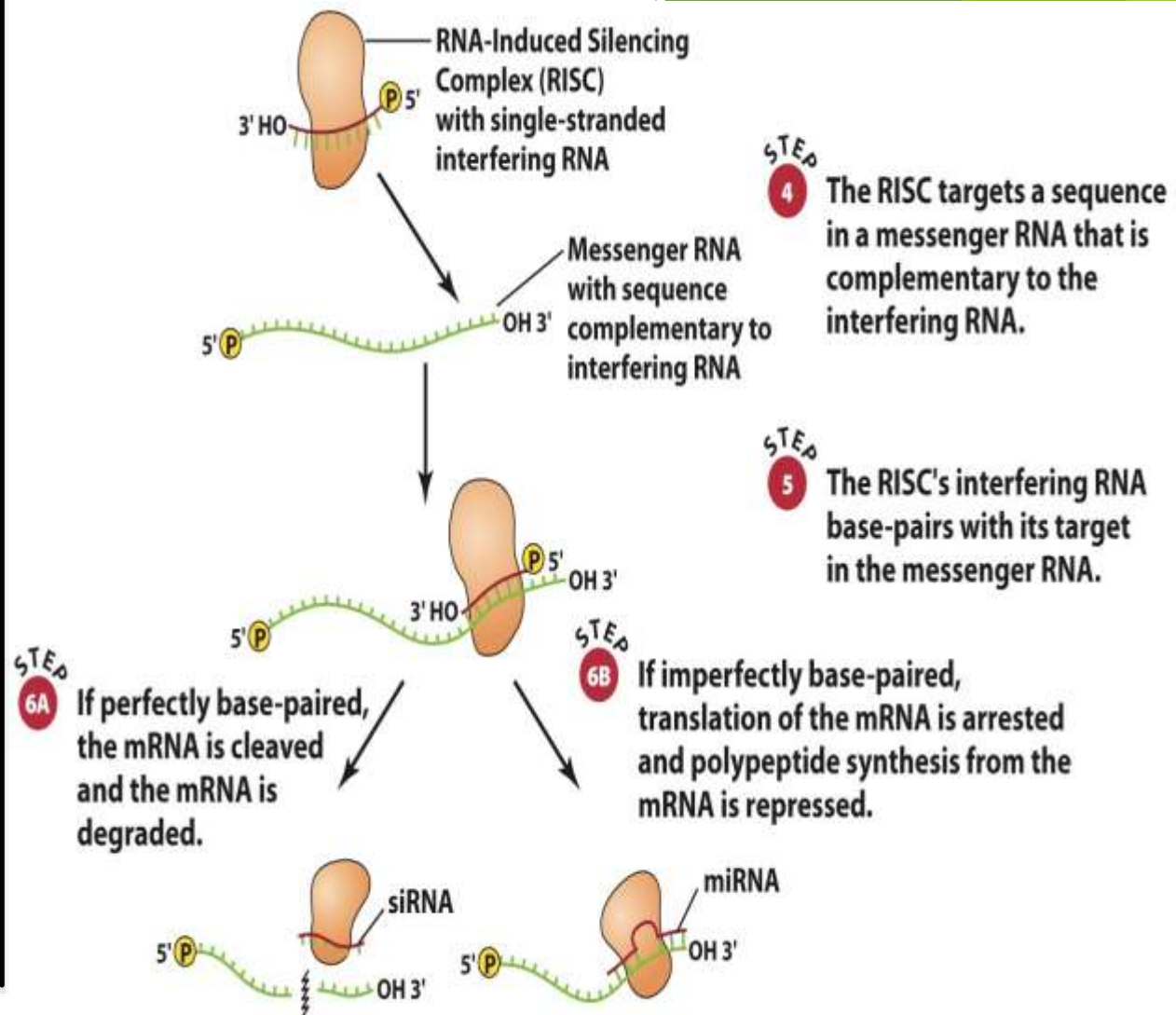




**STEP 1** A large double-stranded RNA molecule is diced into small, double-stranded interfering RNAs 21-28 base pairs long.

**STEP 2** The small interfering RNAs and proteins assemble into ribonucleoprotein particles.

**STEP 3** The small interfering RNA in a ribonucleoprotein particle is unwound to produce an RNA-Induced Silencing Complex (RISC).



**STEP 4** The RISC targets a sequence in a messenger RNA that is complementary to the interfering RNA.

**STEP 5** The RISC's interfering RNA base-pairs with its target in the messenger RNA.

**STEP 6A** If perfectly base-paired, the mRNA is cleaved and the mRNA is degraded.

**STEP 6B** If imperfectly base-paired, translation of the mRNA is arrested and polypeptide synthesis from the mRNA is repressed.

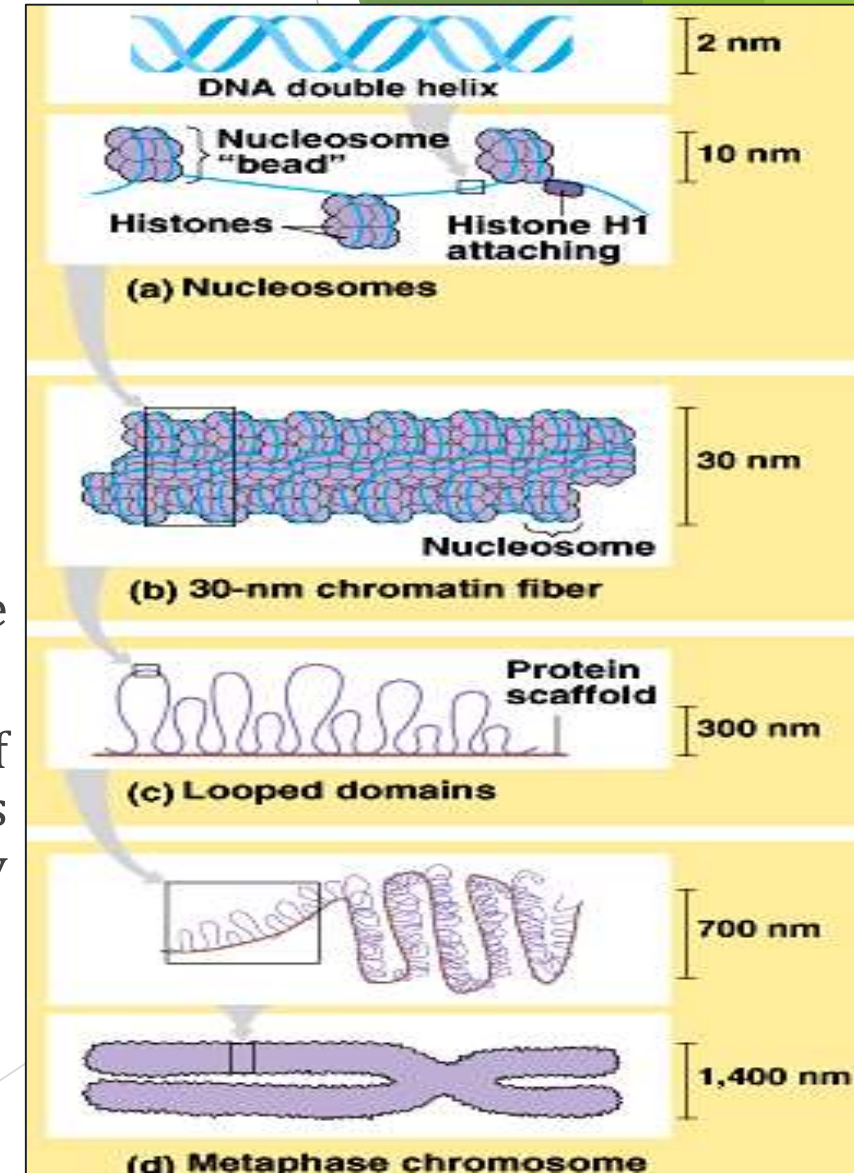
**FIG:** Summary of events involved in RNA interference pathways.

# Sources of siRNA and miRNA

- ▶ miRNAs are derived from endogenous transcripts of the *mir* genes.
- ▶ Long double-stranded RNA that is transfected or injected into a cell can be processed to form an siRNA. This can be used experimentally to knock down expression of a gene.

# Gene Expression and Chromatin Organization

- ▶ DNA coiling & folding
  - Double helix
  - Nucleosomes
  - Chromatin fiber
  - Looped domains
  - Chromosome
- ▶ Various aspects of chromatin organization influence the transcription of genes.
- ▶ Eukaryotic chromosomes are composed of about equal parts of DNA and protein. Collectively, we refer to this material as **chromatin**. The chemical characteristics of chromatin vary along the length of a chromosome.



# Euchromatin and Heterochromatin

- ▶ Variation in the density of chromatin within the nuclei of cells leads to differential staining of sections of chromosomes. The deeply staining material is called **heterochromatin**, and its lightly staining counterpart is called **euchromatin**. A combination of genetic and molecular analyses has shown that the vast majority of eukaryotic genes are located in euchromatin.
- ▶ Moreover, when euchromatic genes are artificially transposed to a heterochromatic environment, they tend to function abnormally, and, in some cases, not to function at all. This impaired ability to function can create a mixture of normal and mutant characteristics in the same individual, a condition referred to as **position-effect variegation**.



# Position-Effect Variegation in *Drosophila*

- ▶ The *white mottled* allele is a good example of position-effect variegation. In this case, a wild-type allele of the *white* gene has been relocated by an inversion, with one break near the euchromatic *white* locus and the other in the basal heterochromatin of the X chromosome. This rearrangement interferes with the normal expression of the *white* gene and causes a mottled-eye phenotype.
- ▶ Apparently, the euchromatic *white* gene cannot function well in a heterochromatic environment.



Courtesy of Joseph Fong, University of Minnesota

**FIG:** The variegated eye color phenotype of *Drosophila* that carry the *white mottled* allele in a rearranged X chromosome.

# Molecular Organization of Transcriptionally Active DNA

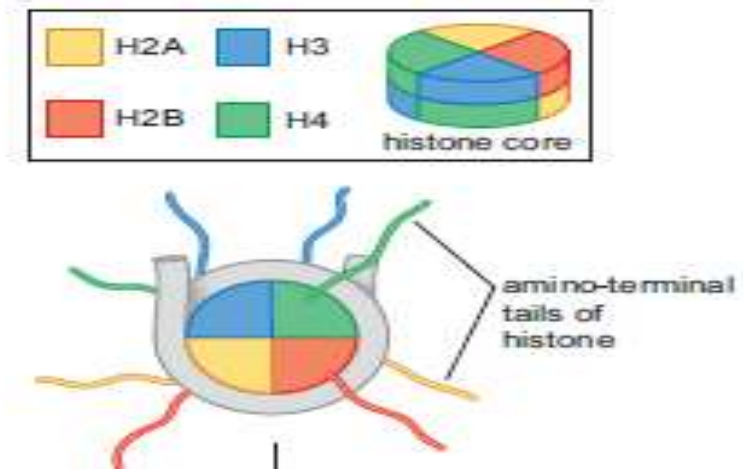
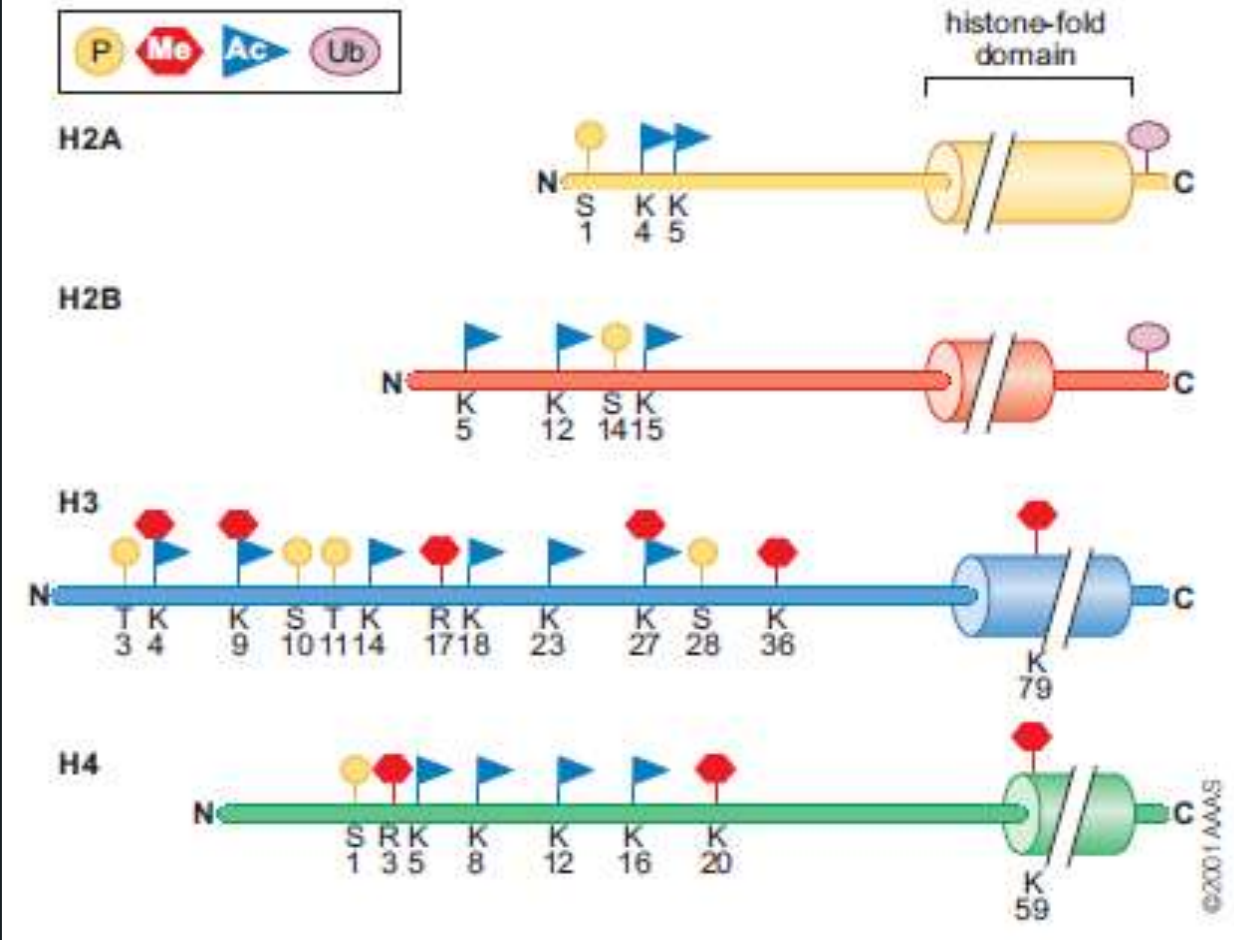
- ▶ Transcriptionally active DNA is more sensitive to DNase I than non-transcribed DNA.
  - ▶ Groudine and Weintraub extracted chromatin from chicken red blood cells and partially digested it with DNase I. Then they probed the residual chromatin material for sequences of two genes,  $\beta$ -globin, which is actively transcribed in red blood cells, and ovalbumin, which is not. They found that over 50 percent of the  $\beta$ -globin DNA had been digested by the DNase I enzyme, compared with only 10 percent of the ovalbumin DNA. These results strongly implied that the actively transcribed gene was more “open” to nuclease attack.
- ▶ The nuclease sensitivity of transcriptionally active genes depends on the presence of two small non-histone proteins, HMG14 and HMG17. When these proteins are removed from active chromatin, nuclease sensitivity is lost; when they are added again, it is restored.
- ▶ The promoter and enhancer regions of active genes contain DNase I hypersensitive sites.
  - ▶ The functional significance of these hypersensitive sites is still unclear, but some evidence suggests that they may mark regions in which the DNA is locally unwound, perhaps because transcription has begun.

# Chromatin Remodeling

- ▶ The presence of nucleosomes and of complexes of histones and DNA provide a barrier against the ready association of transcription factors with specific DNA regions. The disruption of nucleosome structure is therefore an important part of eukaryotic gene regulation.
- ▶ The alteration of nucleosomes, to facilitate the action of RNA polymerase, by multiprotein complexes in preparation for transcription is called **chromatin remodeling**.
- ▶ Chromatin can be modified in two general ways. The first involves changes to nucleosomes, and the second involves modifications to DNA.

# Histone Modifications

- ▶ Histone proteins can be modified by-
  - ▶ Acetylation
  - ▶ Methylation
  - ▶ Phosphorylation
  - ▶ Ubiquitinylation
- ▶ Occur on the tail regions.
- ▶ There are occasional modifications within the histone fold.





# Histone Acetylation

- Acetylation of histones unwinds DNA : Histone acetyltransferase (HAT)
  - Attachment of acetyl groups to histones
    - Conformational change in histone proteins
    - Transcription factors have easier access to genes
  - Loosely wrapped around histones
    - Enables transcription
    - Genes turned on
- Deacetylation
  - Histone Deacetylase (HDATs) : repressing the expression

# Histone Methylations

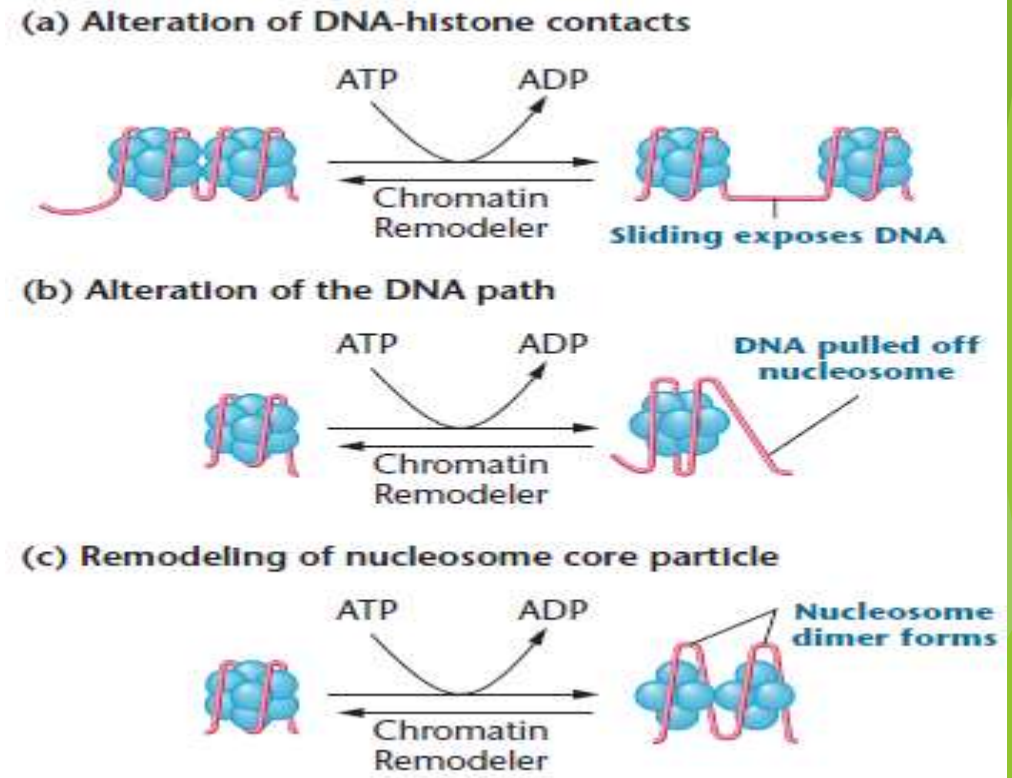
- ▶ Methylation of arginine residue at the 4<sup>th</sup> position of histone H4-
  - ▶ opens the chromatin structure
  - ▶ leading to transcriptional activation
- ▶ Methylation of lysines residues at the 4<sup>th</sup> and 79<sup>th</sup> position of histone H3
  - ▶ opens the chromatin structure
  - ▶ leading to transcriptional activation
- ▶ Methylation of lysines residues at the 9<sup>th</sup> and 27<sup>th</sup> position of histone H3
  - ▶ condenses the chromatin structure
  - ▶ leading to transcriptional inactivation

# Histone Phosphorylation and Ubiquitination

- ▶ Kinases—enzymes that transfer phosphate groups to molecules—may also play a role along with these chromatin-remodeling complexes.
  - ▶ acetylation of lysine-14 in histone H4 is often preceded by phosphorylation of serine-10 in that molecule
  - ▶ opens the chromatin structure
  - ▶ leading to transcriptional activation
- ▶ Ubiquitination
  - ▶ Ubiquitination of H2A – Transcriptional inactivation
  - ▶ Ubiquitination of H2B - Transcriptional activation

# Disruption of Nucleosome Structure by SWI/SNF Complex

- ▶ SWI/SNF complexes may loosen the attachment between histones and DNA, resulting in the nucleosome sliding along the DNA and exposing regulatory regions.
- ▶ Alternatively, they may loosen the DNA strand from the nucleosome core, or they may cause reorganization of the internal nucleosome components.
- ▶ In all cases, the DNA is left transiently exposed to association with transcription factors and RNA polymerase.



**FIG:** Three ways by which chromatin remodelers, such as the SWI/SNF complex, alter the association of nucleosomes with DNA. (a) The DNA-histone contacts may be loosened, allowing the nucleosomes to slide along the DNA, exposing DNA regulatory regions. (b) The path of the DNA around a nucleosome core particle may be altered. (c) Components of the core nucleosome particle may be rearranged, resulting in a modified nucleosome structure.

## How Does Histone Modification Alter Nucleosome Function?

- Acetylation and phosphorylation each acts to reduce the overall positive charge of the histone tails.
- Acetylation of lysine neutralizes its positive charge - this loss of positive charge reduces the affinity of the tails for the negatively charged backbone of the DNA.
- Modification of the histone tails affects the ability of nucleosome arrays to form more repressive higher-order chromatin structure.
- Histone amino-terminal tails are required to form the (30-nm) chromatin fiber, and modification of the tails modulates this function.

# DNA Methylation

- ▶ The DNA of most eukaryotic organisms can be modified after DNA replication by the enzyme-mediated (DNA methyltransferase) addition of methyl groups to bases, a mechanism called **DNA methylation**. DNA methylation most often occurs at position 5 of cytosine (5-methylcytosine) of CG doublets in DNA, usually on both strands:

5'- mCpG - 3'

3'- GpCm - 5'

- ▶ Potentially, methylatable CpG sequences are not randomly distributed throughout the genome, but are concentrated in CpG-rich regions (about 1 to 2 kb long), called **CpG islands**, located at the 5' ends of genes, usually in promoter regions. In the genome of eukaryotic species, approximately 2-7 percent of the cytosine residues are methylated.
- ▶ CpG dinucleotides occur less often than expected in mammalian genomes, probably because they have been mutated into TpG dinucleotides over the course of evolution.

# DNA Methylation

- ▶ Methylation of the cytosine in both strands is termed full methylation, whereas methylation of only one strand is called hemimethylation .
- ▶ Where methylated DNA is found, it is associated with transcriptional repression.
- ▶ **Evidence of a role for methylation in eukaryotic gene expression is based on a number of observations:**
- ▶ First, an inverse relationship exists between the degree of methylation and the degree of expression. Large transcriptionally inert regions of the genome, such as the inactivated X chromosome in mammalian female cells, are often heavily methylated.

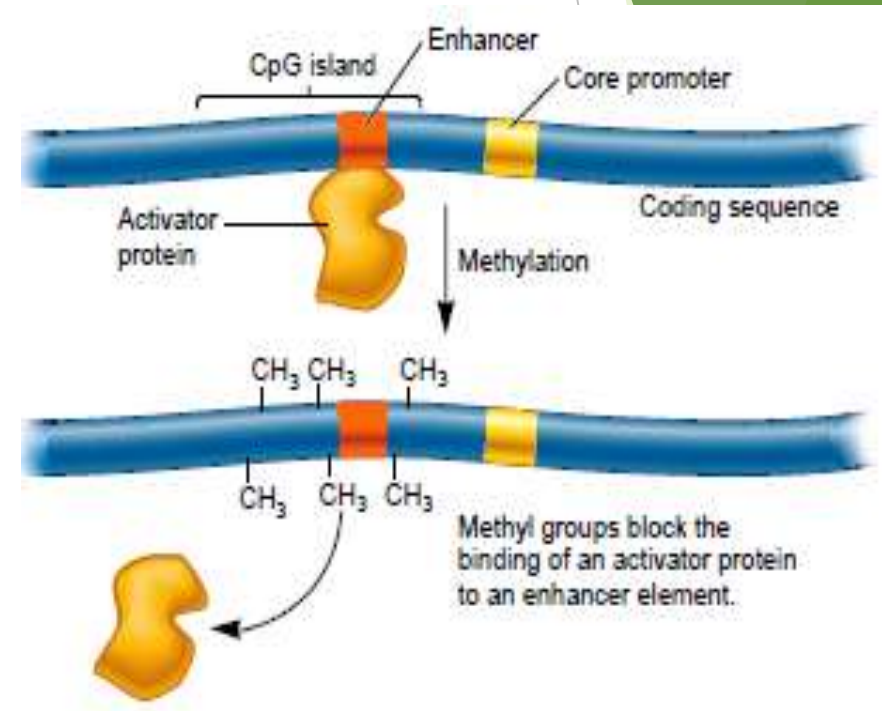


# DNA Methylation

- ▶ Second, methylation patterns are tissue specific and, once established, are heritable for all cells of that tissue. It appears that proper patterns of DNA methylation are essential for normal mammalian development. Despite this heritability, DNA methylation in specific gene regions can be altered by **methylase** and **demethylase** enzymes in order to activate or silence regions of DNA.
- ▶ Perhaps the most direct evidence of a role for methylation in gene expression comes from studies using base analogs. The nucleoside **5-azacytidine** can be incorporated into DNA in place of cytidine during DNA replication. This analog cannot be methylated, causing the undermethylation of the sites where it is incorporated. The incorporation of 5-azacytidine into DNA changes the pattern of gene expression and stimulates expression of alleles on inactivated X chromosomes. In addition, the presence of 5-azacytidine in DNA can induce the expression of genes that would normally be silent in certain differentiated cells.

# How Might DNA Methylation Affect Gene Regulation?

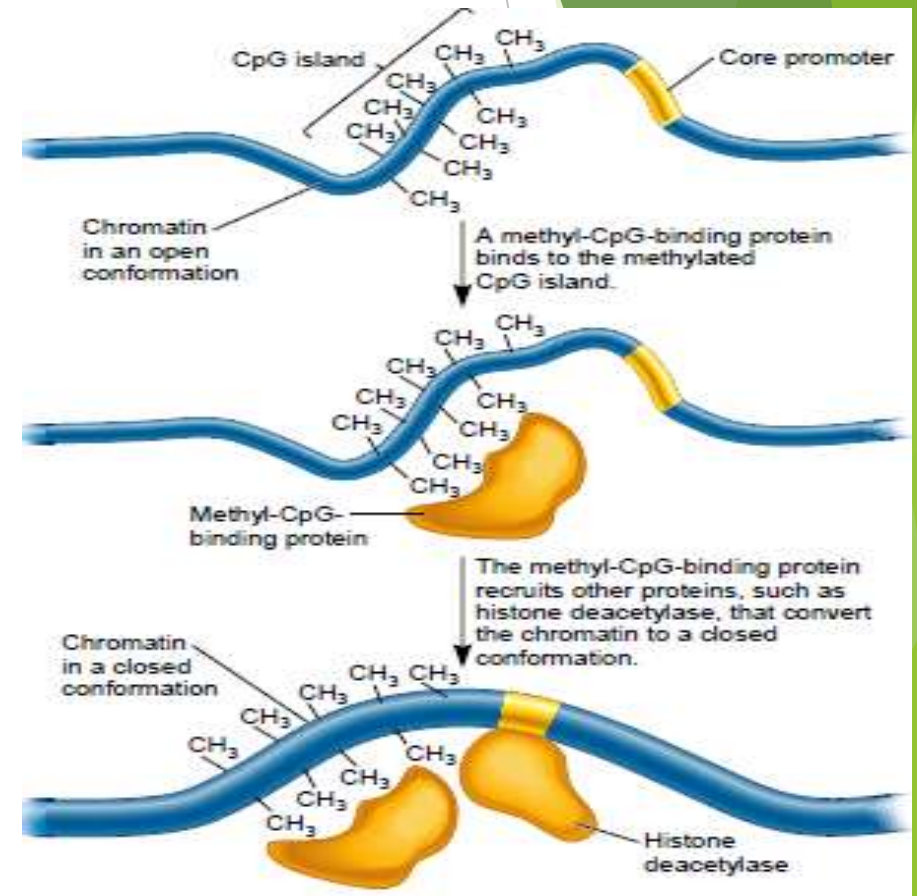
- ▶ **Methylation can affect transcription in two general ways:**
- ▶ First, methylation of CpG islands may prevent or enhance the binding of regulatory transcription factors to the promoter region. For example, methylated CG sequences could prevent the binding of an activator protein to an enhancer element, presumably by the methyl group protruding into the major groove of the DNA. The inability of an activator protein to bind to the DNA would inhibit the initiation of transcription.



**FIG: Transcriptional silencing via methylation.** The methylation of a CpG island may inhibit the binding of transcriptional activators to the promoter region.

# How Might DNA Methylation Affect Gene Regulation?

- ▶ A second way that methylation inhibits transcription is via proteins known as methyl-CpG-binding proteins (MeCPs), which bind methylated sequences. These proteins contain a domain called the methyl-binding domain that specifically recognizes a methylated CG sequence. Once bound to the DNA, the methyl-CpG-binding protein recruits other proteins to the region that inhibit transcription.
- ▶ For example, methyl-CpG binding proteins may recruit histone deacetylase to a methylated CpG island near a promoter. Histone deacetylation removes acetyl groups from the histone proteins, which makes it more difficult for nucleosomes to be removed from the DNA. In this way, deacetylation tends to inhibit transcription.



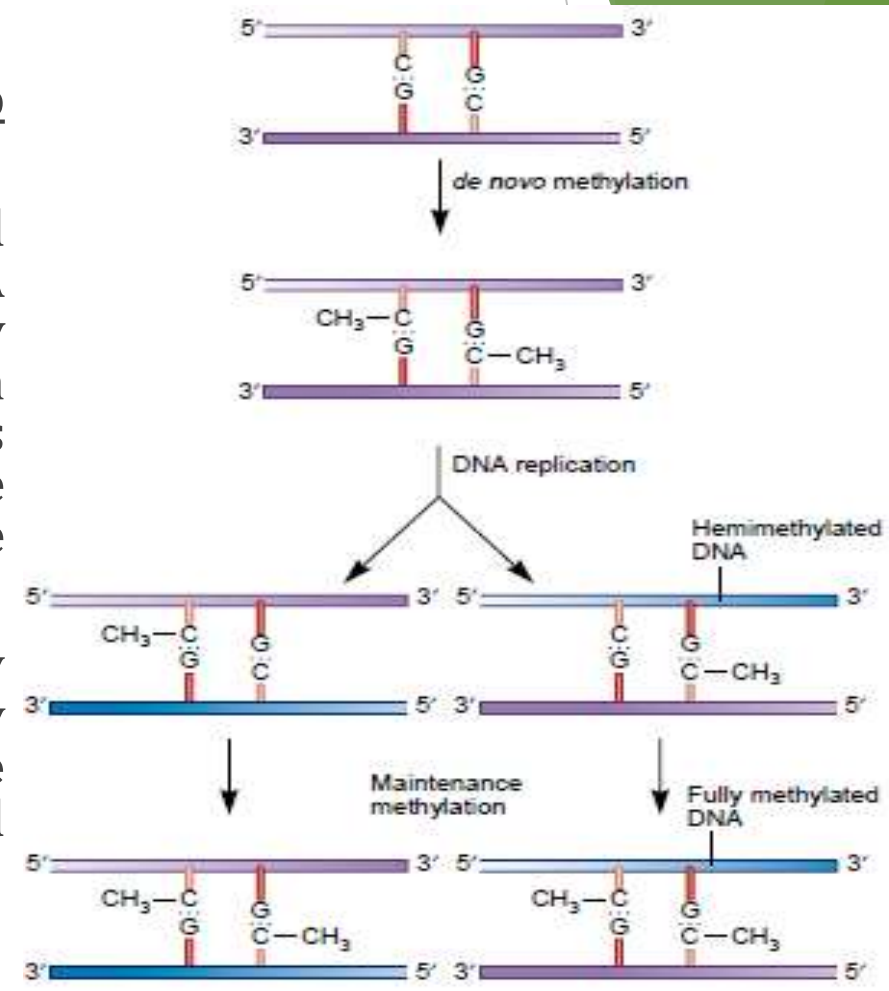
**FIG: Transcriptional silencing via methylation.** The binding of a methyl-CpG-binding protein to a CpG island may lead to the recruitment of proteins, such as histone deacetylase, that convert chromatin to a closed conformation and thus suppress transcription.

# DNA Methylation is Heritable

- ▶ Methylated DNA sequences are inherited during cell division.
- ▶ Experimentally, if fully methylated DNA is introduced into a plant or vertebrate cell, the DNA will remain fully methylated even in subsequently produced daughter cells.
- ▶ However, if the same sequence of nonmethylated DNA is introduced into a cell, it will remain nonmethylated in the daughter cells.
- ▶ These observations indicate that the pattern of methylation is retained following DNA replication and, therefore, is inherited in future daughter cells.

# DNA Methylation is Heritable

- ▶ How can methylation be inherited from cell to cell?
- ▶ The DNA in a particular cell may become methylated by ***de novo* methylation**—the methylation of DNA that was previously unmethylated. When a fully methylated segment of DNA replicates in preparation for cell division, the newly made daughter strands contain unmethylated cytosines. Because only one strand is methylated, such DNA is said to be hemimethylated.
- ▶ This hemimethylated DNA is efficiently recognized by **DNA methyltransferase**, which makes it fully methylated. This process is called **maintenance methylation**, because it preserves the methylated condition in future cells.



**FIG: A molecular model for the inheritance of DNA methylation.** The DNA initially undergoes *de novo* methylation, which is a rare, highly regulated event. Once this occurs, DNA replication produces hemimethylated DNA molecules, which are then fully methylated by DNA methyltransferase. This process, called maintenance methylation, is a routine event that is expected to occur for all hemimethylated DNA.



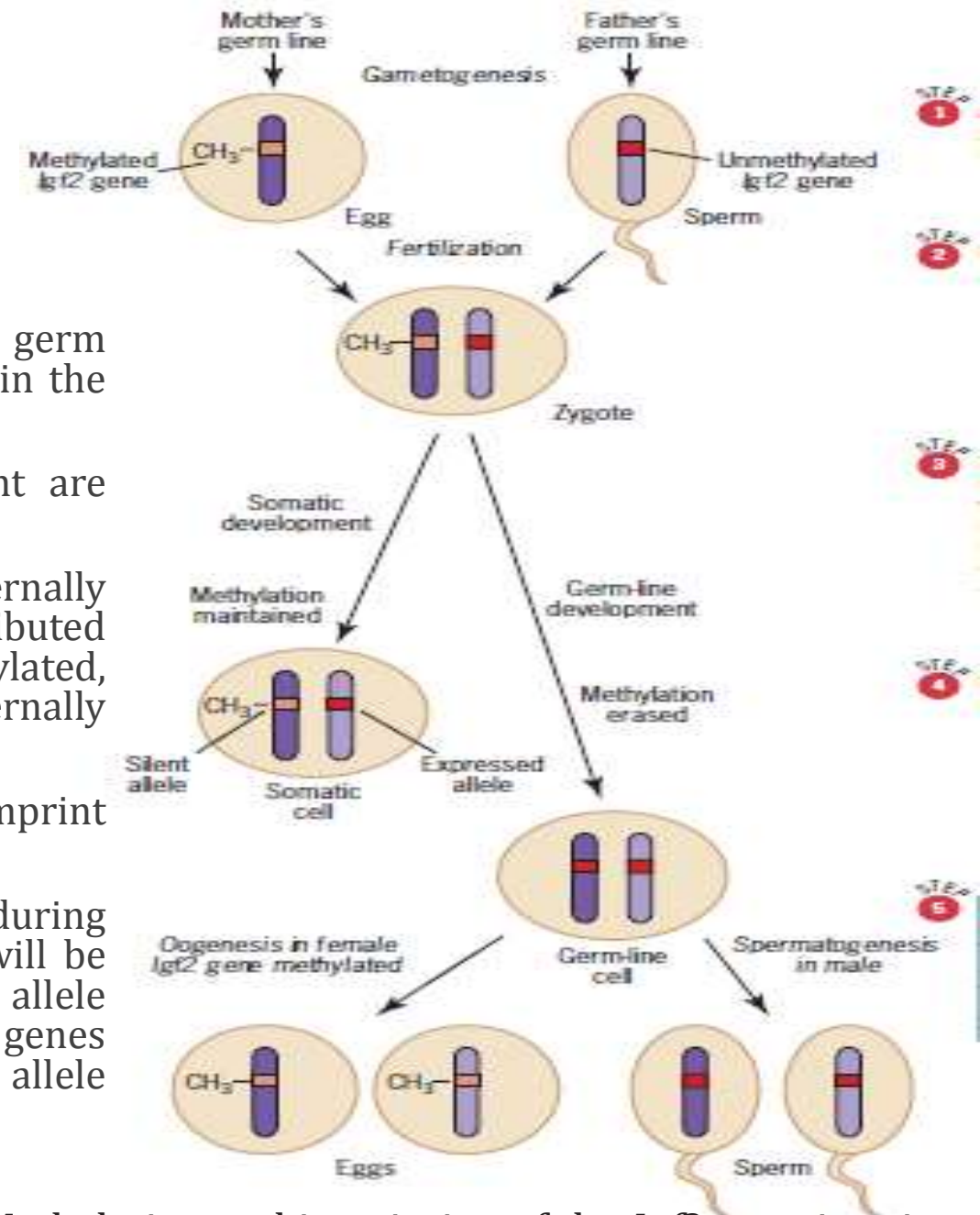
# DNA Imprinting

- ▶ Whenever the expression of a gene is conditioned by its parental origin, geneticists say that the gene has been **imprinted**—a term intended to convey the idea that the gene has been marked in some way so that it “remembers” which parent it came from.
- ▶ For example, in mice, the *Igf 2* gene, which encodes an *insulin-like growth factor*, is expressed when it is inherited from the father but not from the mother. By contrast, a gene known as *H19* is expressed when it is inherited from the mother but not from the father.
- ▶ Recent molecular analysis has demonstrated that the mark that conditions the expression of a gene is methylation of one or more CpG dinucleotides in the gene’s vicinity. These methylated dinucleotides are initially formed in the parental germ line.



# DNA Imprinting

- ▶ **Step 1:** Alleles of the *Igf2* gene are imprinted in the parental germ lines—methylated in the female germ line and not methylated in the male germ line.
- ▶ **Step 2:** Imprinted alleles of the *Igf2* gene from each parent are combined in the zygote at fertilization.
- ▶ **Step 3:** During development of the somatic tissues, the maternally contributed allele remains methylated while the paternally contributed allele remains unmethylated. In somatic cells, only the unmethylated, paternally contributed allele is expressed. The methylated, maternally contributed allele is silent.
- ▶ **Step 4:** During development of the germ line, the methylation imprint is erased.
- ▶ **Step 5:** Methylation is reestablished during oogenesis, but not during spermatogenesis. Thus, if the mouse is female, all *Igf2* genes will be methylated, even if they are copies of the unmethylated *Igf2* allele inherited from the father. If the mouse is male, none of the *Igf2* genes will be methylated even if they are copies of the methylated *Igf2* allele inherited from the mother.



**FIG:** Methylation and imprinting of the *Igf2* gene in mice. The gene is methylated in females but not in males.

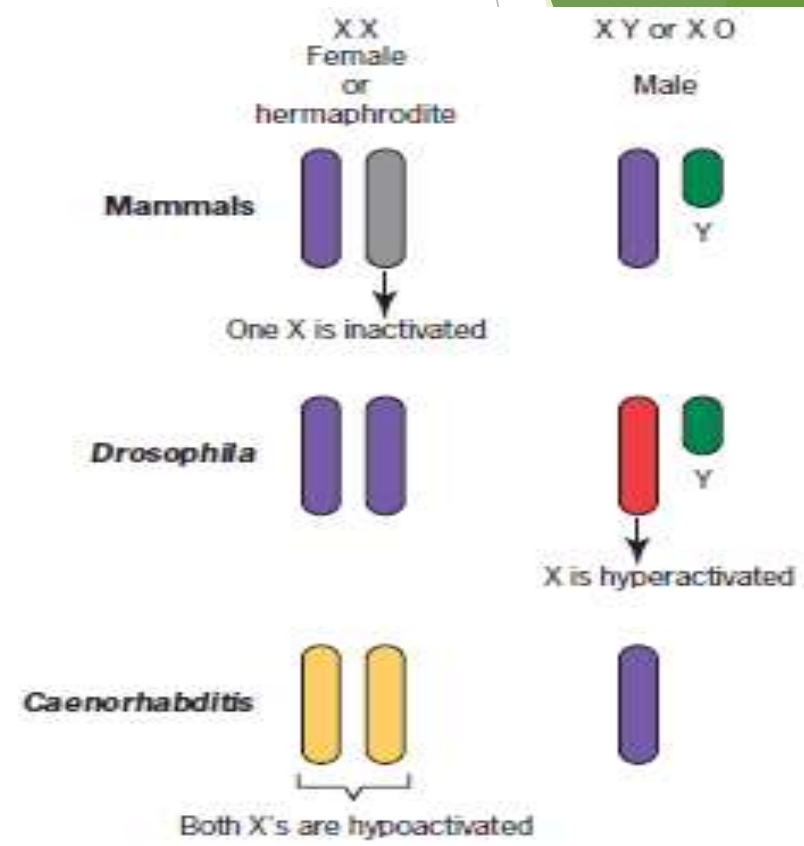
# Activation and Inactivation of Whole Chromosomes

- ▶ Organisms with an XX/XY or XX/XO sex- determination system face the problem of equalizing the activity of X-linked genes in the two sexes.
- ▶ **In mammals:**
  - ▶ Sex determination: XX individuals are female and XY individuals are male.
  - ▶ Dosage compensation: random inactivation of one of the two X chromosomes in female.
- ▶ **In *Drosophila*:**
  - ▶ Sex determination: XX individuals are female and XY individuals are male.
  - ▶ Dosage compensation: neither of the two X chromosomes in a female is inactivated; instead, the genes on the single X chromosome in a male are transcribed more vigorously to bring their output in line with that of the genes on the two X chromosomes in a female.

# Activation and Inactivation of Whole Chromosomes

## ► In nematode *Caenorhabditis elegans*:

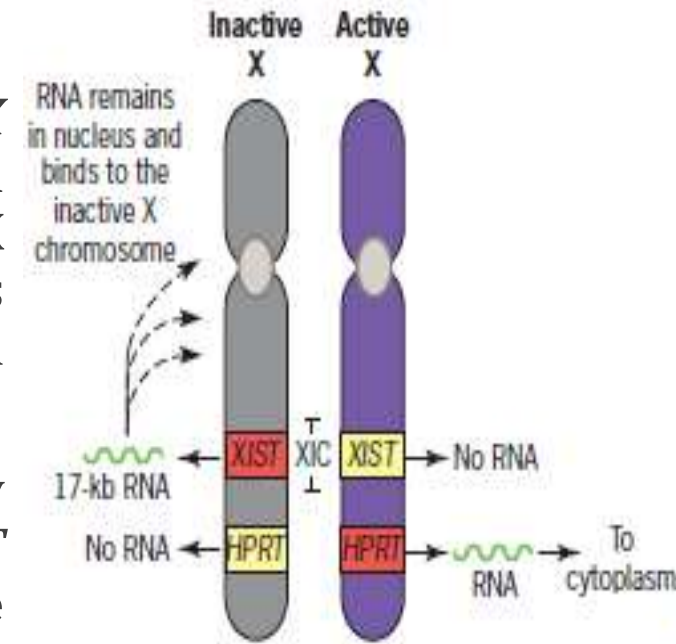
- Sex determination: XX individuals are hermaphrodites (they function as both male and female), and XO individuals are males.
  - Dosage compensation: partial repression of the genes on both of the X chromosomes in the hermaphrodites.
- Therefore, mammals, flies, and worms have solved the problem of X-linked gene dosage in different ways. In mammals, one of the X chromosomes in females is **inactivated**; in *Drosophila*, the single X chromosome in males is **hyperactivated**; and in *C. elegans*, both of the X chromosomes in hermaphrodites are **hypoactivated**.



**FIG:** Three mechanisms of dosage compensation for X-linked genes: inactivation, hyperactivation, and hypoactivation.

# Mechanisms of Dosage Compensation

- ▶ **Inactivation of X chromosome in mammals:**
- ▶ X chromosome inactivation begins at a particular site called the *X inactivation center (XIC)* and then spreads in opposite directions toward the ends of the chromosome. Curiously, not all genes on an inactivated X chromosome are transcriptionally silent. One that remains active is called *XIST* (for *X inactive specific transcript*); this gene is located within the XIC.
- ▶ In human beings the *XIST* gene encodes a 17-kb transcript devoid of any significant open reading frames. It therefore seems unlikely that the *XIST* gene codes for a protein. Instead, the RNA itself is probably the functional product of the *XIST* gene. Though polyadenylated, this RNA is restricted to the nucleus and is specifically localized to inactivated X chromosomes; it does not appear to be associated with active X chromosomes in either males or females.



**FIG:** Expression of the *XIST* gene in the inactive X chromosome of human females. For comparison, the expression of the *HPRT* gene on the active X chromosome is shown. This gene encodes hypoxanthine phosphoribosyl transferase, an enzyme that plays a role in the metabolism of purines.

# Mechanisms of Dosage Compensation

- ▶ **Hyperactivation of X chromosome in *Drosophila*:**
- ▶ In *Drosophila*, dosage compensation requires the protein products of at least five different genes. Null mutations in these genes result in male-specific lethality because the single X chromosome in males is not hyperactivated. These dosage compensation genes are therefore called male-specific lethal (*msl*) loci, and their products are called the MSL proteins.
- ▶ The binding of the MSL proteins to the male's X chromosome is facilitated by two types of RNA molecules called *roX1* and *roX2* (for RNA on the X chromosome) that are transcribed from genes on the X chromosome.
- ▶ The MSL proteins form a complex that is joined by the *roX* RNAs and this complex then binds to 30 to 40 sites along the male's X chromosome, including the loci that contain the two *roX* genes. From each of these entry sites, the MSL/*roX* complex spreads bidirectionally until it reaches all the genes on the male's X chromosome that need to be hyperactivated.



# Mechanisms of Dosage Compensation

- ▶ The process of hyperactivation may involve chromatin remodeling by the MSL/*roX* complex. One of the MSL proteins is a histone acetyl transferase, and a particular acetylated version of histone H4 is exclusively associated with hyperactivated X chromosomes.
- ▶ **Hyporactivation of X chromosome in *Caenorhabditis elegans*:**
- ▶ In *C. elegans*, dosage compensation involves the partial repression of X-linked genes in the somatic cells of hermaphrodites.
- ▶ Like the MSL proteins in *Drosophila*, the proteins encoded by several specific genes of *C. elegans* bind specifically to the X chromosomes and partially repress the transcription of their genes.
- ▶ They bind only when two X chromosomes are present. The proteins apparently do not bind to the single X chromosome in males, nor do they bind to any of the autosomes in either males or hermaphrodites.



The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, ranging from light lime to dark forest green. These shapes are concentrated on the right side of the image, creating a dynamic, layered effect. The left side of the image is a solid, light gray background.

*THANKS To ALL*