# 14 - Gene Regulation in Eukaryotes

## Comparing Gene Regulation in Prokaryotes and Eukaryotes

The *lac* operon provides an excellent example of how bacteria perform gene regulation in response to an environment that lacks glucose yet contains lactose. In the case of the *lac* operon, we learned that gene regulation involves an activator protein (CAP) and a *lac* repressor protein. The effector molecules cAMP and allolactose regulate CAP and the *lac* repressor binding to regulatory DNA sequences (CAP site, operator) near the promoter for the *lac* operon. Ultimately the binding of the CAP and the *lac* repressor proteins determine if the sigma (σ) factor protein and the RNA polymerase core enzyme activate transcription.

Even though gene regulation in prokaryotes and eukaryotes is similar (e.g., both involve activator proteins, repressor proteins, effector molecules, and regulatory DNA sequences), eukaryotic gene regulation is more complex. This complexity is needed to produce multicellular eukaryotic organisms with cells in each tissue having unique phenotypes. For example, a white blood cell (leukocyte) and a muscle cell have the same collection of structural genes; however, gene regulation ensures that a leukocyte expresses leukocyte-specific proteins, while a muscle cell expresses muscle-specific proteins. Further, many eukaryotic organisms progress from a fertilized egg through complex developmental stages to produce the mature adult. Gene regulation ensures that embryonic genes are expressed only during embryonic development, while other genes are expressed solely in the adult.

Regulation of a typical eukaryotic gene involves **combinatorial control**. For example, a single eukaryotic gene can be regulated by a combination of:

- **Activator** proteins binding to **enhancer** DNA sequences**.**
- **Repressor** proteins binding to **silencer** DNA sequences**.**
- **Regulation of activator and repressor protein function.** This regulation of activator and repressor proteins involves effector molecules, covalent modification, and protein-protein interactions.
- **Modifying the structure of chromatin to activate or repress transcription.** Modifying chromatin involves chemically modifying histone proteins or altering the arrangement of nucleosomes near the core promoter of a gene.
- **DNA methylation to silence transcription.** The methylation of cytosine bases near the core promoter region of a gene inhibits transcription.

### Key Questions

- What is meant by combinatorial control?
- What combinations of processes influence the transcription of a eukaryotic gene?

## Core Promoter vs. Regulatory Promoter

We learned in Part 9 that transcription in eukaryotes involves several types of DNA sequences. The **core promoter**, for example, determines where RNA polymerase II will bind to the DNA and begin transcription**.** The core promoter includes the **TATA box (-25 sequence),** which serves as the binding site for the general transcription factor protein **TFIID** and the **+1 site**, the first base in the template DNA strand that is transcribed by RNA polymerase II. For transcription to occur, the TATA box and the +1 site must be present. If these two sequences are the only sequences present upstream of a gene, the gene will be transcribed at a low, yet constant rate (the so-called **basal** level of transcription).

In addition to the core promoter, many eukaryotic genes include a **regulatory promoter** (see **figure 14.1**)**.** The components of the regulatory promoter are required for transcription levels higher than the basal level provided by the core promoter. A common regulatory promoter component that is present in many eukaryotic genes is the **CAAT box**. The CAAT box is located at -80 and has the sequence 5'-GGCCAATCT-3'. Another common regulatory promoter component is a **GC box** (5'- GGGCGG – 3') located at -100. The CAAT and GC boxes are the binding sites for certain **activator** proteins. Thus, the CAAT and GC boxes can be considered enhancers adjacent to many eukaryotic structural genes.

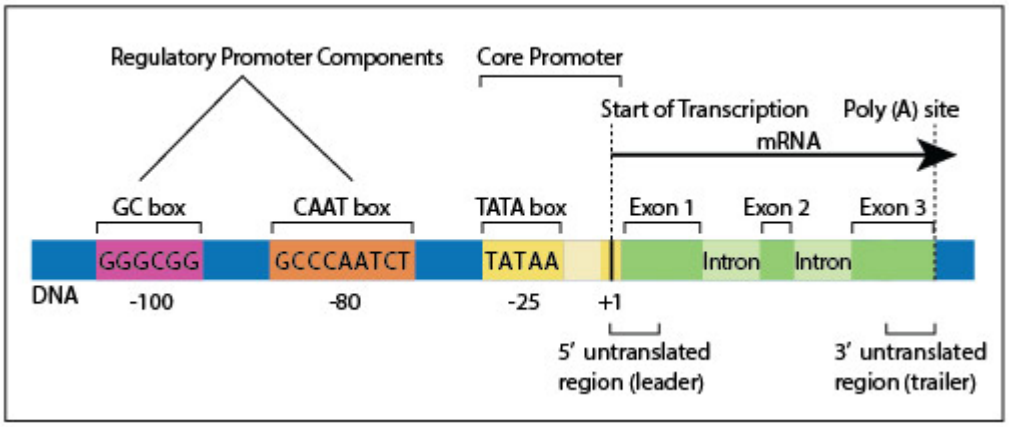



*Figure 14.1* ***Core and Regulatory Promoter*** *--- Image created by SL*

### Key Questions

- What is meant by basal transcription?
- What is the function of the regulatory promoter?
- What are the names of two common DNA sequences found in the regulatory promoters of eukaryotic genes?

## General and Regulatory Transcription Factors

Eukaryotic **transcription factors** are proteins that influence the ability of RNA polymerase II to bind to a eukaryotic core promoter. There are two categories of transcription factor proteins:

- **General transcription factor proteins (GTFs).** The general transcription factor proteins include the **TFIID, TFIIA, TFIIB, TFIIF, TFIIE**, and **TFIIH** proteins described in Part 9. These proteins function to recruit RNA polymerase II to the core promoter to begin transcription. The general transcription factors are required for all transcription events. If these general transcription factors are the only proteins involved, the gene is transcribed at the basal level. The general transcription factors are also required for transcription rates above this basal level.
- **Regulatory transcription factor proteins**. Regulatory transcription factor proteins function to regulate transcription by either increasing transcription above the basal level or decreasing transcription below the basal level. An **activator** protein increases the level of transcription above the basal level; a **repressor** protein decreases the level of transcription below the basal level. Many activator and repressor proteins are only expressed in certain tissues or at certain times during development, thus playing a critical role in tissue-specific or time-specific gene expression.

Transcription factors proteins are ***trans*-acting factors** (i.e., can regulate genes found throughout the genome) and bind to DNA sequences called ***cis*-acting elements** (i.e., the DNA binding sites near the controlled gene) (see **figure 14.2**). However, these *cis*-acting elements are not always adjacent to the core promoter. Some *cis*-acting elements are within the gene that they control or can be thousands of base pairs away.

Recall that the **mediator** protein complex communicates the signals from activator and repressor proteins to RNA polymerase II. Mediator thus serves as a link between regulatory transcription factors, the GTF proteins, and RNA polymerase II, thereby determining the overall rate of transcription.

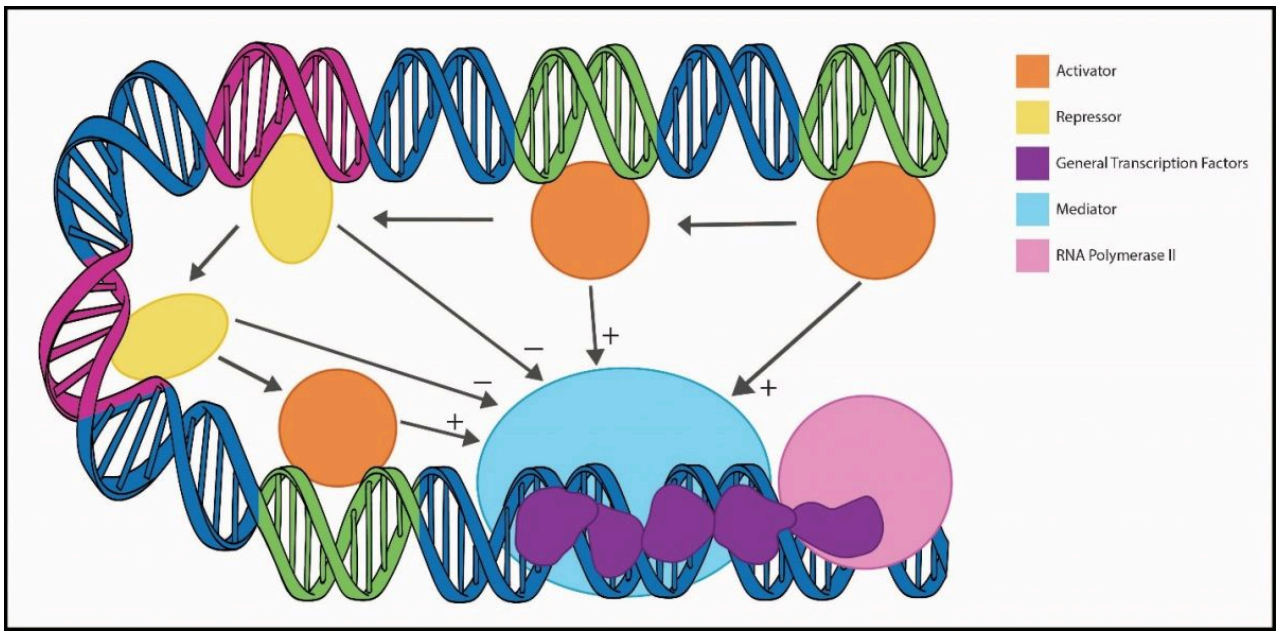



*Figure 14.2* ***Trans-acting factors binding to cis-acting elements****. In this case, mediator interprets three activation signals and two silencing signals. Overall, transcription is increased above the basal level. --- Image created by SL.*

## Key Questions

- Review the functions of TFIID, TFIIH, and mediator.
- Which transcription components are considered *trans*-acting factors?
- Which transcription components are considered *cis*-acting elements?

## Enhancers and Silencers

Other regulatory DNA sequences assist the core promoter and regulatory promoter to regulate transcription by serving as the binding sites for transcription factor proteins. The binding of regulatory transcription factors to these DNA sequences may:

- **Increase the rate of transcription.** Transcription can increase 1000-fold when **activator** proteins bind to **enhancer** DNA sequences (**up-regulation**). Activator proteins and enhancer DNA sequences are generally responsible for tissue-specific expression of a gene.
- **Decrease the rate of transcription.** Transcription can decrease below the basal level when **repressor** proteins bind to **silencer** DNA sequences (**down-regulation**). Repressor proteins and silencer DNA sequences are generally responsible for tissue-specific repression of a gene.

A particular gene can be regulated by transcription factor proteins bound to different combinations of enhancer and silencer DNA sequences (see **figure 14.2**). The combination of the transcription factor proteins and regulatory DNA sequences involved determines the transcription pattern of the gene.

### Key Questions

- Review the functions of activator proteins, repressor proteins, enhancer DNA sequences, and silencer DNA sequences (see Part 9).

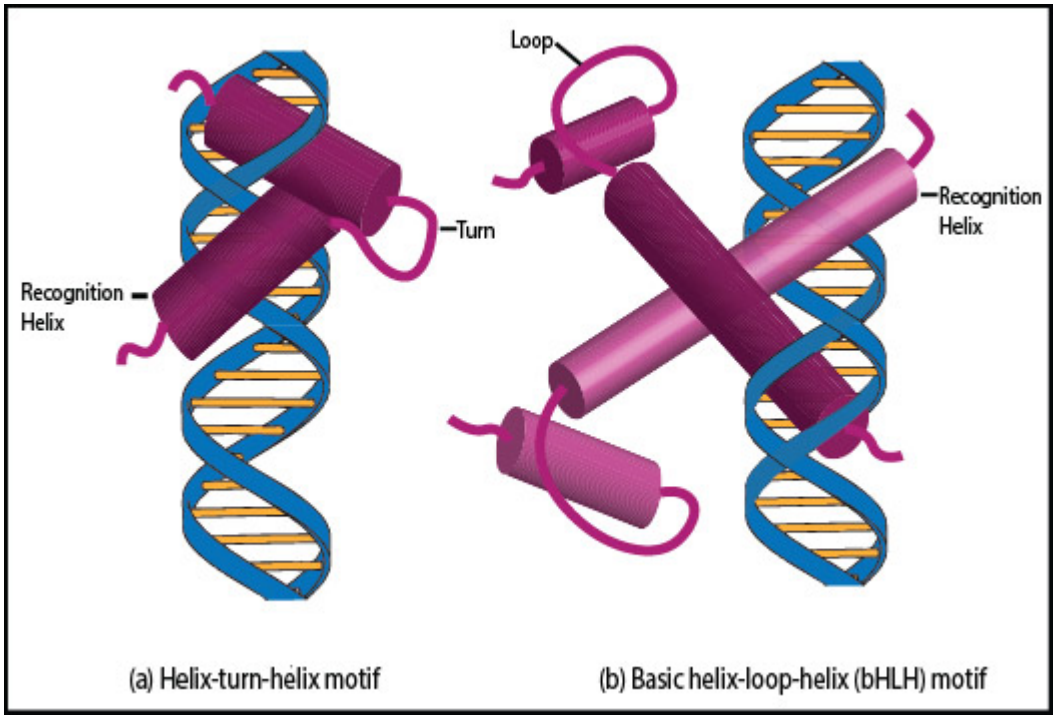
## Structural Features of Transcription Factors

Transcription factor proteins have been identified in many organisms, including viruses, bacteria, fungi, plants, and animals. Nearly all transcription factor proteins contain conserved structural features that are important in either binding to regulatory DNA sequences, effector molecules, or other transcription factor proteins. For example, most transcription factor proteins contain **α-helices**, a type of protein secondary structure. An α-helix is produced when certain amino acids in the polypeptide sequence interact through hydrogen bonding to produce a helical structure. Importantly, the α-helix is the proper width to bind to the major groove in DNA. Thus, the α-helix is often used by transcription factors proteins to recognize specific base pair sequences located in the major groove of the DNA.

Four common **structural motifs** are found in transcription factor proteins. These structural motifs, based upon the α-helix structure described above, include (see **figure 14.3**):

- **Helix-turn-helix (HTH) motif**. The HTH motif includes two α-helices separated by a "turn" of 3-4 amino acids. One α-helix is called the **recognition helix**, and functions to bind to a specific base pair sequence in the DNA major groove. This recognition helix also includes basic (positively charged) amino acids that bind to the negatively charged DNA backbone. The helix-turn-helix motif is found in both prokaryotic and eukaryotic transcription factor proteins. For example, many of the transcription factor proteins that we have discussed contain the HTH motif including **sigma (σ) factor**, the ***lac* repressor** protein, and the **catabolite activator protein (CAP)**.
- **Basic helix-loop-helix (bHLH) motif**. The bHLH motif is similar to the helix-turn-helix motif, containing a recognition helix that binds to the DNA major groove. However, instead of a turn, bHLH transcription factors have an amino acid loop to connect two α-helices. bHLH transcription factors play an important role in cell division and differentiation. For example, the **MyoD** and **c-myc** proteins are transcription factor proteins that contain the bHLH motif. The MyoD protein activates muscle-specific genes, while the c-myc protein activates genes involved in cell division.
- **Zinc finger motif**. The zinc finger motif is composed of a finger-like structure composed of an α-helix (i.e., the recognition helix) and two β-strands (another type of protein secondary structure). Electrostatic interactions between zinc ions ($Zn^{2+}$) and negatively charged amino acid side chains within the transcription factor protein stabilize the zinc finger motif. Steroid hormone receptors, including the **glucocorticoid receptor** protein (see below), **testosterone receptor** protein, and the **estrogen receptor** protein contain zinc finger motifs.
- **Leucine zipper motif**. The leucine zipper motif not only contains a recognition helix, but also contains a second α-helix with many hydrophobic leucine amino acids. When the leucine-rich α-helices of two leucine zipper transcription factors interact, they form a **coiled-coil** to exclude water. The coiled-coil resembles a zipper with interlocking leucine amino acids. The DNA sequence is bound by recognition helices that extend from the coiled-coil region of these two transcription factor proteins. The **CREB** protein (see below) contains a leucine zipper motif.

It is important to note that all four transcription factor motif structures described above permit transcription factor proteins to bind to each other. Two identical transcription factor proteins interact to form a transcription factor **homodimer,** or two different transcription factor proteins interact to form a **heterodimer**. For example, both the CAP protein and the *lac* repressor proteins are homodimers, composed of two identical transcription factor protein with HTH motifs. Higher order interactions (trimers, tetramers) are also possible when transcription factor proteins bind to each other.




Caption

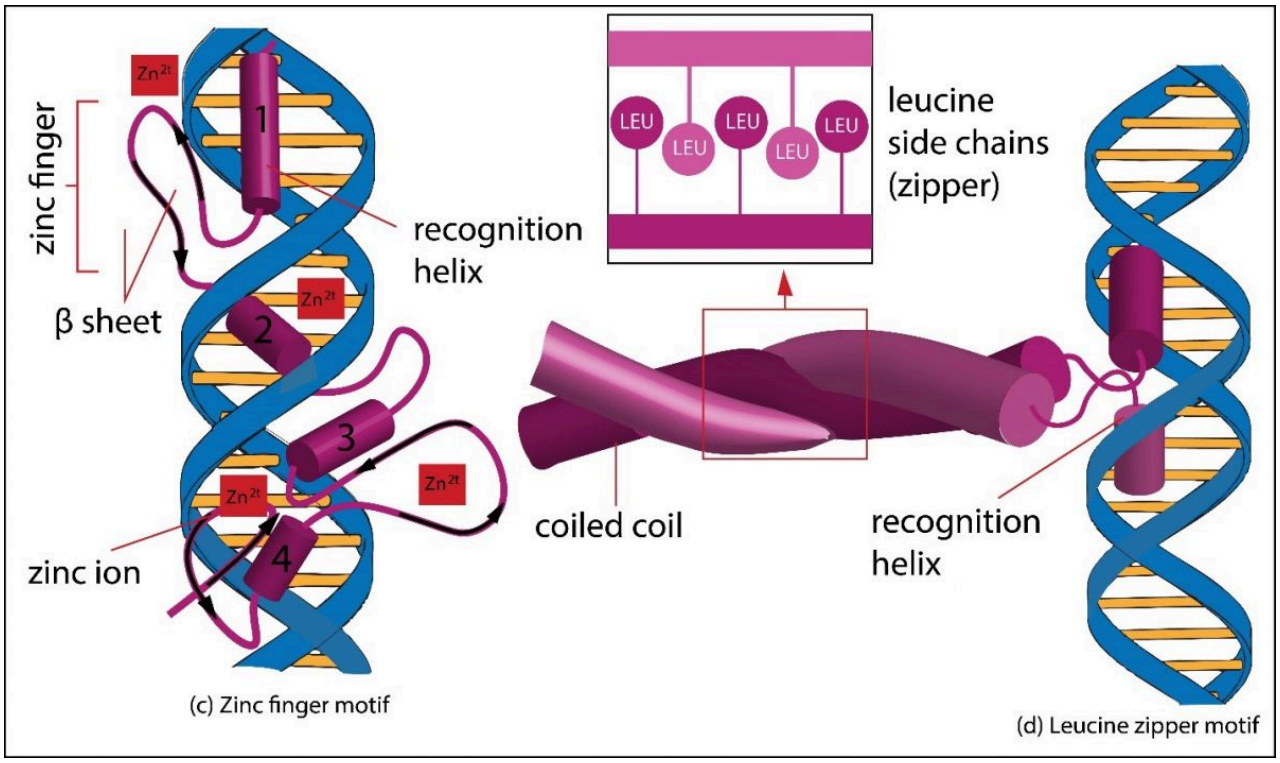



*Figure 14.3* ***Transcription Factor Structural Motifs*** *a) Helix-turn-helix motif b) Basic helix-loop-helix motif c) Zinc finger motif d) Leucine zipper motif --- Images created by SL*

### Key Questions

- What are three examples of transcription factor proteins that contain the helix-turn-helix (HTH) motif?
- What are two examples of transcription factor proteins that contains the basic helix-loop-helix (bHLH) motif?
- What are three examples of transcription factor proteins that contains the zinc finger motif?
- What is an example of a transcription factor protein that contains the leucine zipper motif?
- What protein secondary structure is found in all transcription factor structural motifs?
- What is meant by a transcription factor homodimer or heterodimer?

## Mechanisms to Regulate Transcription Factor Proteins

If an activator protein is present in a cell, it does not always bind to an enhancer DNA sequence and up-regulate transcription. Similarly, a repressor protein does not always bind to a silencer DNA sequence and repress transcription. The DNA-binding activities of activator and repressor proteins are regulated in three ways:

- **Effector binding.** Small effector molecules bind to transcription factor proteins, change the conformation (shape) of the transcription factor, and influence the ability of the transcription factor protein to bind to enhancer or silencer DNA sequences. In animals, steroid hormones such as **glucocorticoid**, **testosterone**, and **estrogen** are effector molecules that regulate the functions of transcription factor proteins.
- **Transcription factor dimerization.** The formation of transcription factor homodimers or heterodimers influences binding to enhancer or silencer DNA sequences.
- **Covalent modification.** The addition of phosphate groups **(phosphorylation)** to activator or repressor proteins can stimulate binding to enhancer or silencer DNA sequences.

Note that for a particular gene, one or more of the above mechanisms may be involved in regulating gene expression. For example, the glucocorticoid receptor transcription factor protein (see below) is regulated by effector binding and dimerization, while the CREB transcription factor protein is regulated by dimerization and covalent modification.

### Key Questions

- Describe the three ways that activator and repressor proteins can be regulated.
- What are three examples of eukaryotic effector molecules?

## Regulating Transcription Through TFIID

Regulatory transcription factor proteins (activator and repressor proteins) influence the ability of RNA polymerase II to transcribe a gene. However, these regulatory transcription factor proteins do not typically bind to RNA polymerase II directly. Instead, transcription factor proteins communicate DNA binding indirectly to RNA polymerase II through other protein complexes. Eukaryotic regulatory transcription factors influence RNA polymerase II activity through **TFIID**, **mediator**, the enzymes involved in **chromatin remodeling,** and the enzymes involved in **DNA methylation**.

Consider first the regulation of RNA polymerase II through the TFIID protein. Recall that TFIID is the general transcription factor protein that binds to the TATA box (the -25 sequence) within the core promoter. TFIID recruits the other five general transcription factors (TFIIA, TFIIB, TFIIF, TFIIH, and TFIIE) that bring RNA polymerase II to the +1 site and activate RNA polymerase II to begin transcription. Suppose an activator protein binds to an enhancer DNA

sequence (see **figure 14.4**). This activator protein then encourages TFIID to bind to the TATA box, and TFIID then recruits the other general transcription factors and RNA polymerase II to the +1 site. As a result, transcription is up-regulated. Suppose instead that a repressor protein binds to a silencer DNA sequence adjacent to a gene. The repressor protein then prevents TFIID from binding to the TATA box. The absence of TFIID on the core promoter prevents the other general transcription factors and RNA polymerase II from binding to the core promoter. As a result, transcription is down-regulated.

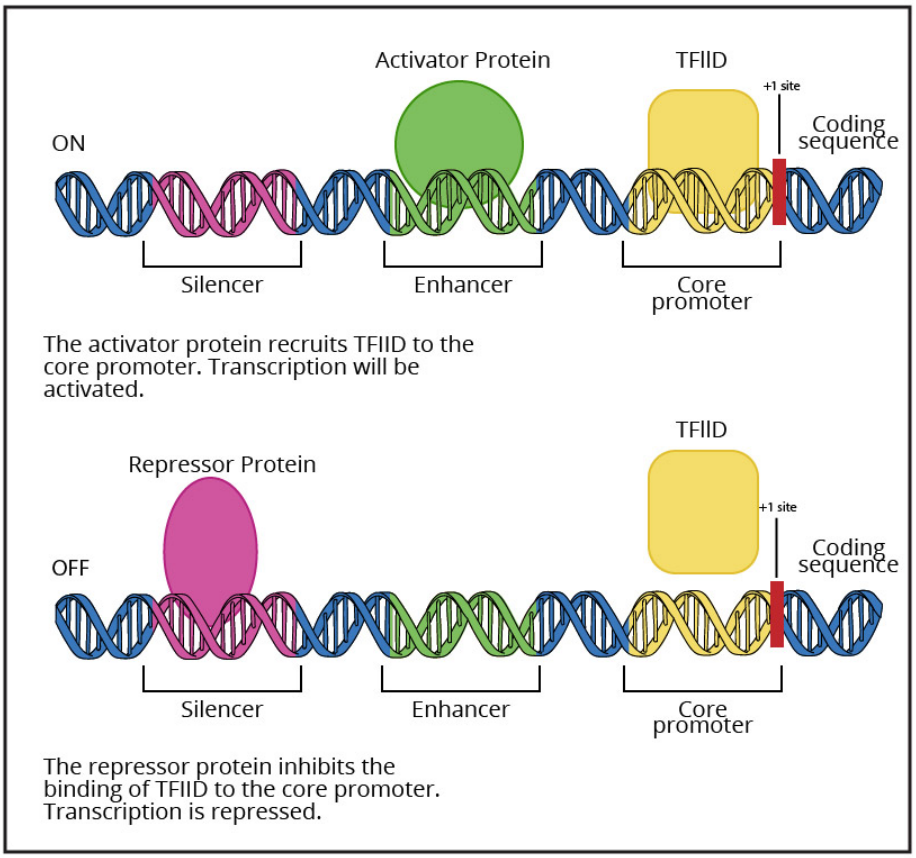



*Figure 14.4* ***Regulating TFIID*** *- Image created by SL*

## Key Questions

- How do activator and repressor proteins influence TFIID?

# Regulating Transcription Through Mediator

**Mediator** is a protein complex that mediates the interaction between the regulatory transcription factors (i.e., activator and repressor proteins) and RNA polymerase II. If mediator activates RNA polymerase II, transcription begins. Suppose an activator protein binds to an enhancer DNA sequence (see **figure 14.5**). The activator protein in turn activates mediator, and mediator then activates the general transcription factor protein **TFIIH**. Next, TFIIH acts as a helicase to separate the template and coding DNA strands. TFIIH also acts as a kinase, phosphorylating RNA polymerase II to begin transcription.

Suppose a repressor protein binds to a silencer DNA sequence instead. The repressor protein then inhibits the activity of mediator. As a result, mediator fails to activate TFIIH, and TFIIH fails to separate the template and coding DNA

strands.  TFIIH also fails to phosphorylate RNA polymerase II, preventing the initiation of transcription.  Note that the DNA between the enhancer/silencer DNA sequences and the core promoter can form a loop to permit the proteins described above to bind to each other.

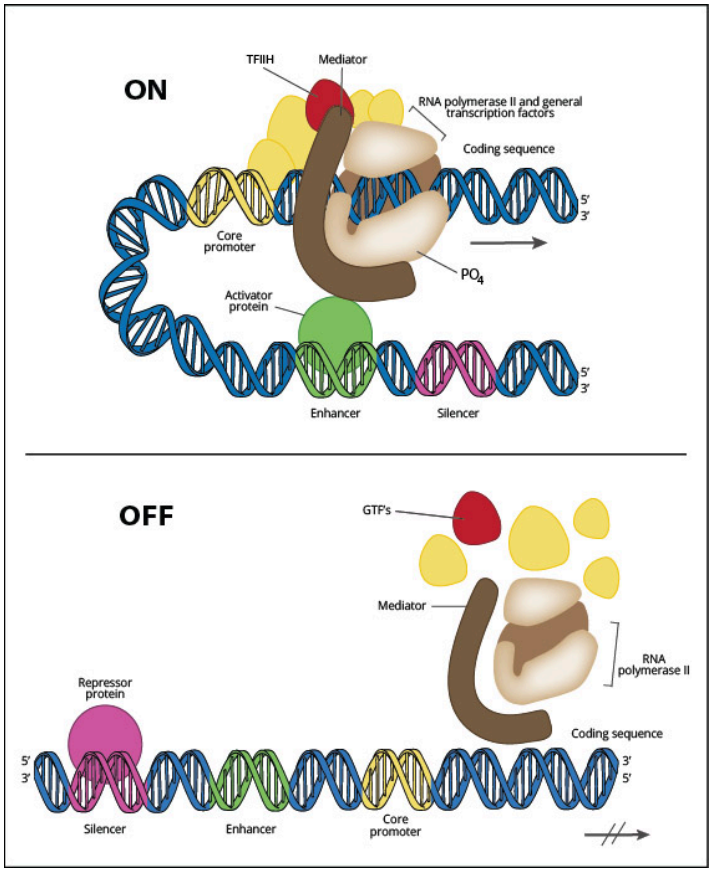



*Figure 14.5* ***Regulating Mediator*** *--- Image created by SL*

## Key Questions

- How do activator and repressor proteins influence the activity of mediator?

# Transcription Activation Using the Glucocorticoid Receptor

Now let's apply what we have learned so far to two examples of gene regulation in the human body.  The first example shows how steroid hormones produced by endocrine glands activate the transcription of genes.  For example, **glucocorticoid hormones (GCs)** are released by the adrenal glands in response to fasting, as well as physical activity. The GCs lead to an increase in glucose synthesis, an increase in protein metabolism, an increase in fat metabolism, and a decrease in inflammation.

Glucocorticoid hormones increase the transcription of a gene above the basal level as follows (see **figure 14.6**):

1. The glucocorticoids are steroid hormones, which are nonpolar in structure. As a result, the glucocorticoids cross the cytoplasmic membrane and enter the cytoplasm of a target cell.
2. Glucocorticoids act as effector molecules by binding to an activator protein called **glucocorticoid receptor** that is found in many cell types. Prior to glucocorticoid binding, the glucocorticoid receptor is bound to **HSP90** proteins. HSP90 helps maintain the proper three-dimensional shape of the glucocorticoid receptor, so that glucocorticoid receptor can bind to glucocorticoid hormones produced by the adrenal glands. HSP90 is released when glucocorticoid hormone binds to glucocorticoid receptor.
3. Glucocorticoid binding changes the conformation (shape) of the glucocorticoid receptor, exposing a **nuclear localization signal (NLS).** The NLS is a polypeptide sequence that helps to target the glucocorticoid receptor (with bound glucocorticoid hormone) to the nucleus of the cell.
4. Two glucocorticoid receptors with bound glucocorticoid hormones form a homodimer in the cytoplasm of the cell.
5. The glucocorticoid receptor:glucocorticoid homodimer complex travels to the nucleus of the cell.
6. The glucocorticoid receptor:glucocorticoid homodimer complex binds to two adjacent enhancer DNA sequences called **glucocorticoid response elements (GREs)**. GREs are common enhancers found adjacent to many genes involved in metabolism.
7. The glucocorticoid receptor: glucocorticoid homodimer complex bound to the GRE sequences activates transcription.

Other steroid hormones, such as estrogen and testosterone, are effector molecules that activate transcription by binding to similar cytoplasmic transcription factor proteins. For example, estrogen binds to estrogen receptor proteins to activate transcription, while testosterone binds to testosterone receptor proteins to activate transcription. Both the estrogen receptor and testosterone receptor proteins are regulated by dimerization.

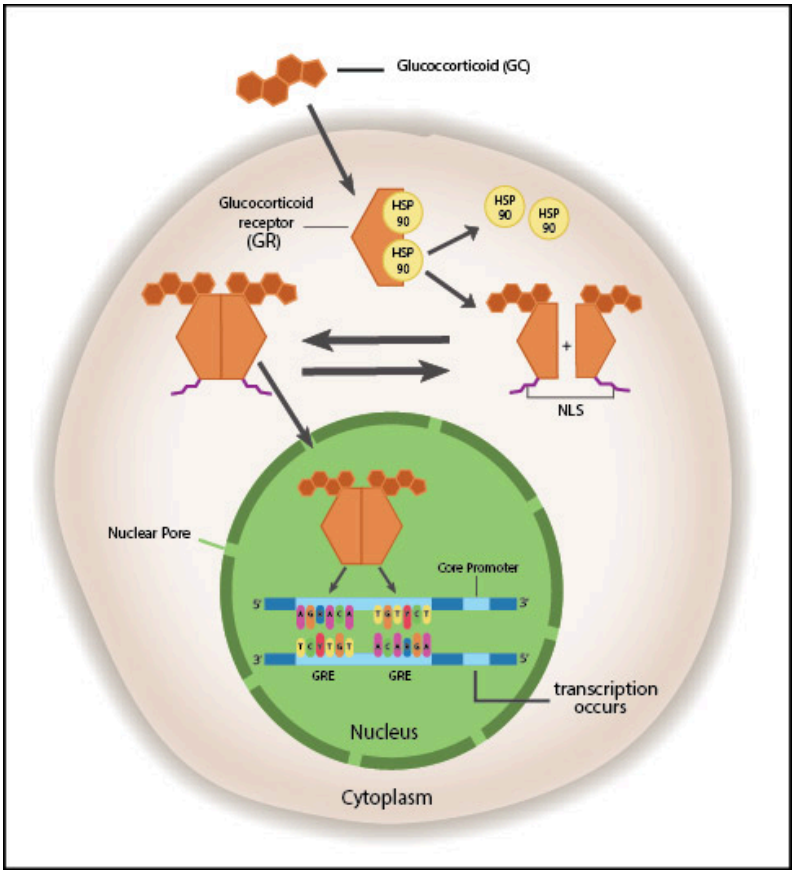



*Figure 14.6* ***Transcription Regulation by Glucocorticoid*** *--- Image created by SL*

## Key Questions

- How does the production of glucocorticoid by an adrenal gland lead to transcriptional activation of a target gene?

# Transcription Activation via CREB

Unlike glucocorticoid, many signaling molecules in the body, such as peptide hormones, growth factor proteins, and cytokine proteins, are not able to diffuse through the cytoplasmic membrane into the cytoplasm of the target cell. Instead, these signaling proteins bind to cell receptors on the surface of a target cell, and the receptor binding signal is then transmitted to the nucleus to activate transcription. Our second example of gene regulation demonstrates how transcription is up-regulated when receptor binding activates the transcription factor protein **cAMP response element-binding protein (CREB)**. Transcription activation via CREB occurs when (see **figure 14.7**):

1. A receptor protein embedded in the cytoplasmic membrane binds to a peptide hormone, growth factor, or cytokine protein.
2. Receptor binding activates a **G protein**.
3. The G protein activates **adenylyl cyclase** inside the cell, which converts ATP into **cAMP**.
4. cAMP binds to and activates **protein kinase A (PKA).**
5. PKA moves into the nucleus and phosphorylates an inactive CREB protein homodimer.
6. The phosphorylated CREB protein homodimer binds to two adjacent enhancer sequences called **cAMP response elements (CREs).**
7. The phosphorylated CREB homodimer bound to the CRE sequences activates transcription.

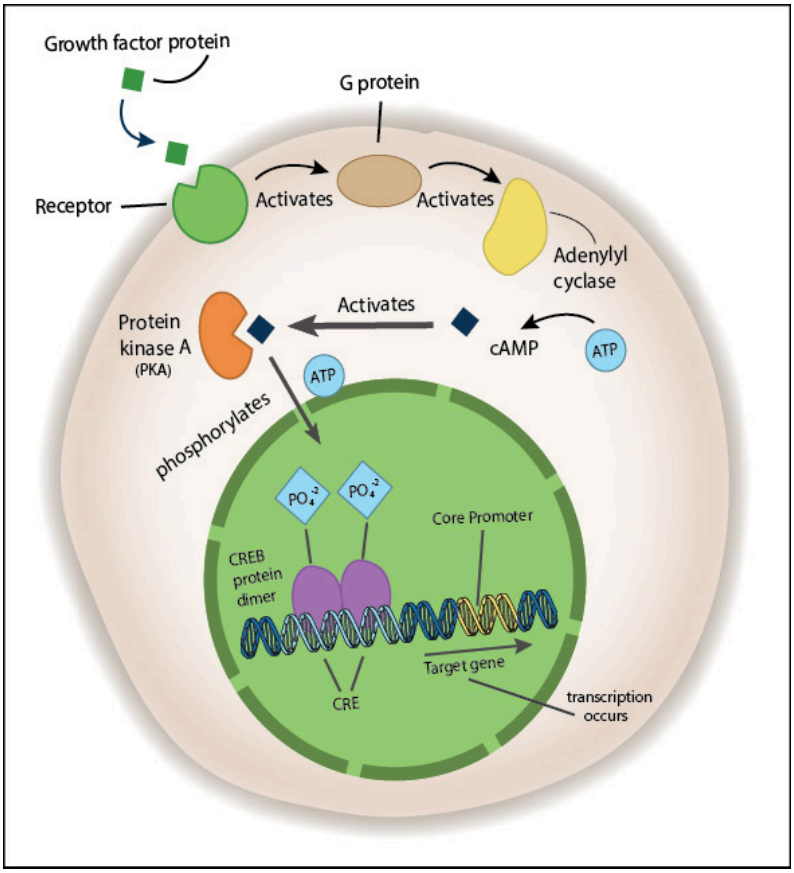



*Figure 14.7* ***Transcriptional Regulation by CREB*** *--- Image created by SL*

## Key Questions

- What is CREB and CRE?
- How does the binding of a signaling protein to a receptor lead to transcriptional activation of a target gene via the CREB pathway?

# Chromosome Compaction and Transcription

The arrangement of nucleosomes on the DNA can also influence the transcription of a nearby gene (for a review of nucleosomes, refer to Part 2). For a gene to be transcribed, RNA polymerase II must be able to bind to the core promoter. If the core promoter region of a gene contains tightly packed nucleosomes (**heterochromatin**), RNA polymerase II struggles to find the core promoter. As a result, the heterochromatin form of DNA is said to be in a **closed conformation** and transcription is limited. Regions of the chromosome with loosely packed or absent nucleosomes are called **euchromatin** (**open conformation**). RNA polymerase II can better access a core promoter located in euchromatin, and as a result, transcription occurs more readily.

Recall that chromatin is a dynamic structure with a specific gene alternating between the closed (heterochromatin) and open (euchromatin) conformations depending on the needs of the cell. When an activator protein binds to an enhancer DNA sequence, chromatin is converted to the open conformation. When a repressor protein binds to a silencer DNA sequence, chromatin is converted to the closed conformation.

## Key Questions

- Review the structure of a nucleosome and the terms heterochromatin and euchromatin (see Part 2).
- What is the difference between the open conformation and the closed conformation?

# Arrangement of Chromatin at the β-globin Gene

As an example of how chromatin structure can influence the transcription of a gene, consider the human β-globin gene (see **figure 14.8**). The β-globin gene, which encodes the β-globin protein components of hemoglobin, is not normally expressed in many cell types, including fibroblast cells. When the DNA region that encompasses the β-globin gene from fibroblasts was analyzed with respect to nucleosomes, scientists discovered that nucleosomes were found at approximately 200 base pairs (bp) intervals from the -3000 to +1500 region of the gene. Note that this closed conformation region from -3000 to +1500 includes the regulatory promoter, core promoter, and the beginning portion of the β-globin gene. This heterochromatin arrangement of nucleosomes makes the β-globin promoter inaccessible to the general transcription factors (GTFs) and RNA polymerase II. As a result, the β-globin gene is not transcribed in fibroblasts.

The β-globin gene is expressed in erythroblasts (precursor red blood cells). When the nucleosome arrangement surrounding the β-globin gene was examined in erythroblasts, a different result was observed. Nucleosomes were displaced from the -500 to +200 region of the β-globin gene. This open conformation (euchromatin) area includes the regulatory promoter, core promoter, and the beginning portion of the β-globin gene. Thus, the GTFs and RNA polymerase II can access the regulatory and core promoter region in erythroblasts, leading to the transcription of the β-globin gene.

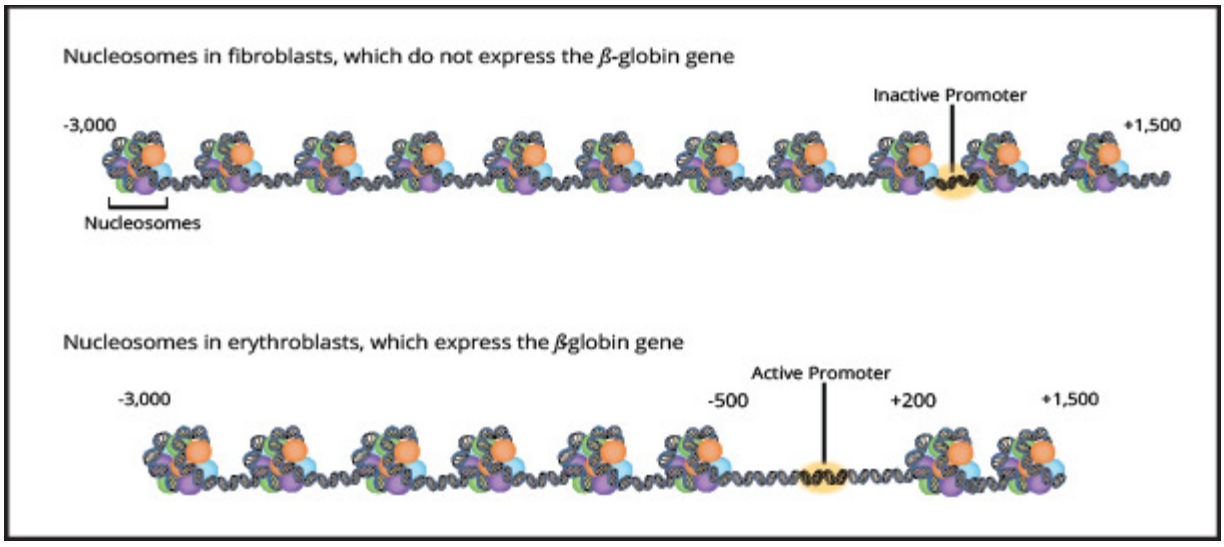



*Figure 14.8* ***Nucleosome arrangement on the B-globin gene*** *--- Image created by SL*

## Key Questions

- In terms of the regulatory and core promoter for the β-globin gene, describe the difference between chromatin structure in fibroblasts and erythroblasts.

# Histone Acetylation

The results from fibroblasts and erythroblasts discussed above suggest that nucleosomes can be altered to influence transcription. Alterations in chromatin structure to promote transcription include the **covalent modification** of histone proteins and the rearrangement of nucleosomes within the promoter region by **ATP-dependent chromatin remodeling** (see **figure 14.9**).

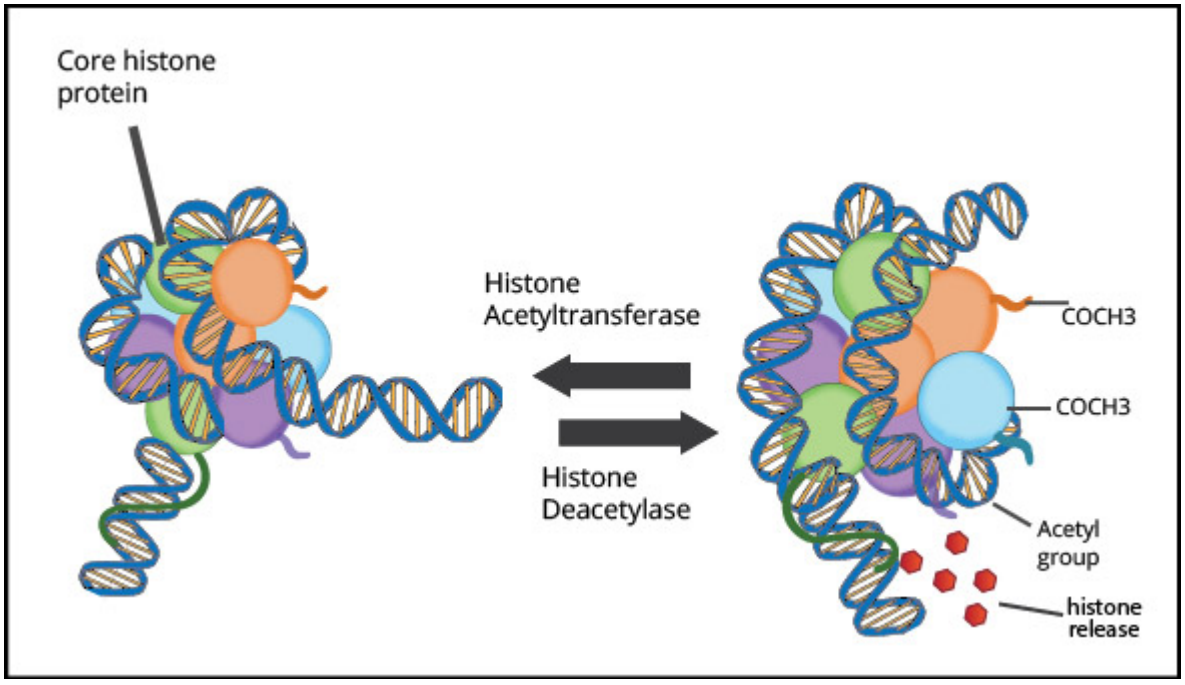
Covalent modification includes the **acetylation** of histone proteins within nucleosomes. Enzymes called **histone acetyltransferases (HATs)** add acetyl chemical groups to the tail regions within histone proteins (refer to the Part 2 reading for a description of histone structure). Acetylation neutralizes the positive charge on lysine amino acids within the histone tail, disrupting the interaction between the histone tail and the negatively charged DNA backbone. As a result, neutralization of the positive charges on the histone tails causes the histones to release from the DNA; the DNA is now more accessible for transcription. When transcription needs to be turned off, the histones are modified by **histone deacetylase (HDAC)** proteins that remove the acetyl groups from histones, restoring the positive charge on the histone tail. The histone tails once again bind to the negatively charged DNA backbone, and the chromatin is converted from the open to the closed conformation (heterochromatin), decreasing transcription of the gene.

Note that when an activator protein binds to an enhancer DNA sequence, the activator recruits HATs to the promoter, activating transcription. Alternatively, when repressor proteins bind to silencer DNA sequences, HDACs are recruited to the promoter, silencing transcription.

# ATP-dependent Chromatin Remodeling

The ATP-dependent chromatin remodeling process uses the energy in ATP to alter the spacing of the nucleosomes in the promoter region near a gene (see figure 14.9). One example of an ATP-dependent chromatin remodeling enzyme is the multi-subunit **SWI/SNF** protein complex. The SWI/SNF protein complex performs at least two types of chromatin remodeling:

- SWI/SNF changes the distribution of nucleosomes along the DNA, creating large gaps between adjacent nucleosomes. When these larger gaps between nucleosomes includes the core promoter region of a gene, transcription in activated.
- SWI/SNF can replace the standard histone proteins (H2A, H2B, H3, and H4; see Part 2) within a nucleosome with **histone variant proteins.** The presence of these histone variant proteins within the modified nucleosome increases transcription.



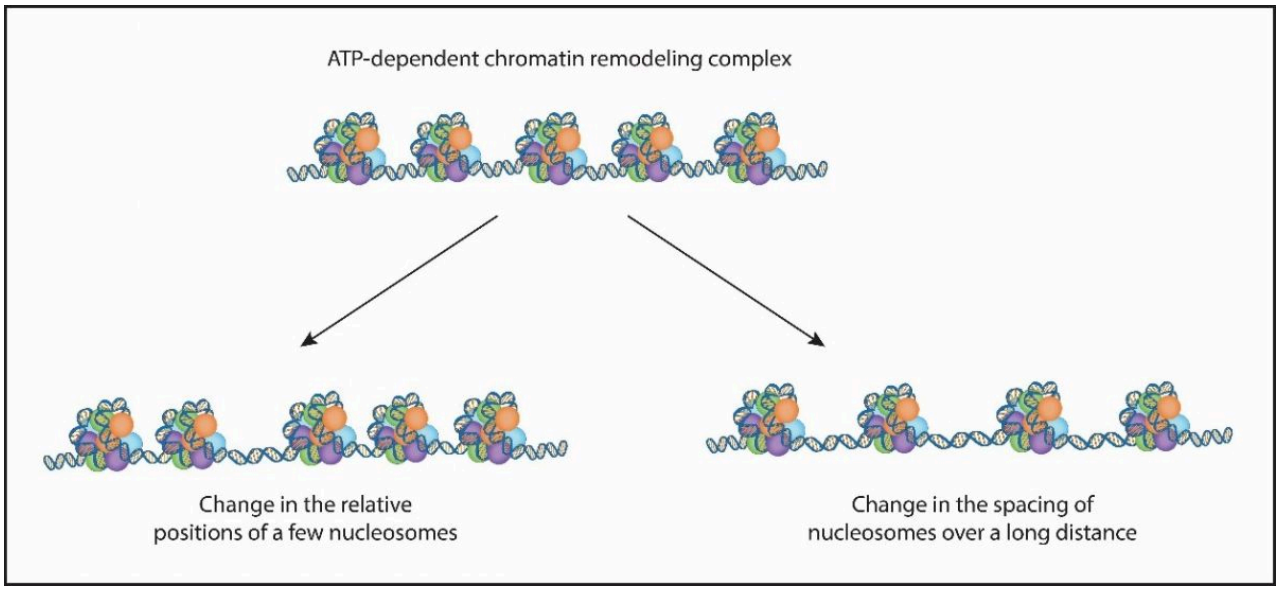




*Figure 14.9* ***Histone Acetylation and ATP-Dependent Chromatin Remodeling*** *--- Images created by SL.*

## Key Questions

- When a HAT is active, what effect does this have on transcription?
- When a HDAC is active, what effect does this have on transcription?
- What is the function of the SWI/SNF protein complex?

# Overview of DNA Methylation

Silencing of gene expression in many eukaryotes involves the **methylation** of DNA sequences near the core promoters of genes. The methyl groups that are added to the DNA double helix block the major groove of the DNA, preventing the

recognition helices (see above) within activator proteins to enhancer sequences from binding to the DNA. Cytosine bases within CG-rich sequences called **CpG islands** are typically targets for DNA methylation. Not surprisingly, many CpG islands are located near the core promoters of genes (see **figure 14.10**). Typical CpG islands are 1,000 – 2,000 base pair (bp) long sequences that contain multiple **CpG sites** (i.e., many 5'-CG-3' dinucleotide sequences in a row). Within CpG islands, adding methyl groups to the cytosine bases on both DNA strands is called **full methylation**. Full methylation inhibits transcription.

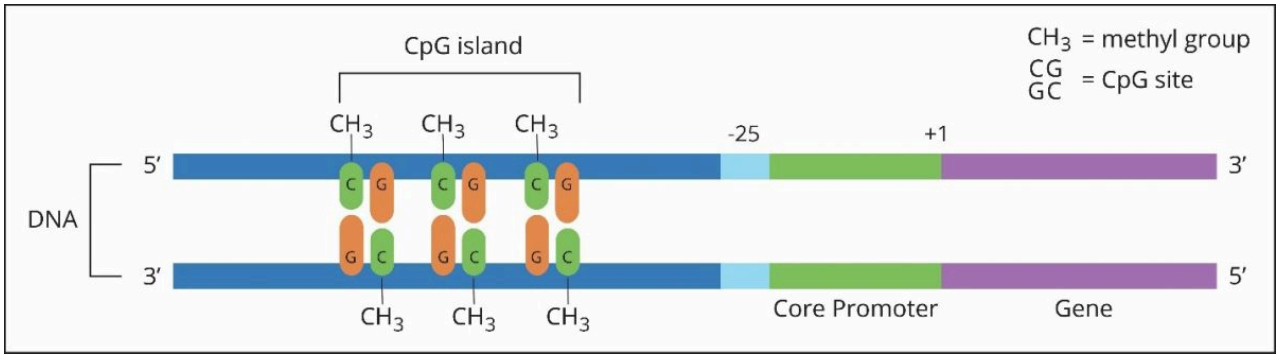



*Figure 14.10* ***Overview of DNA Methylation.*** *CpG islands are the targets for DNA methylation to silence a gene. --- Image created by SL*

**Housekeeping genes** encode proteins that are required for the maintenance of a cell. For example, the structural genes that produce the enzymes involved in glycolysis are housekeeping genes The promoters of these housekeeping genes are typically unmethylated and as a result, housekeeping genes are always transcribed. **Tissue-specific genes** are only expressed in certain cell types. In cell types in which these genes are not expressed, the CpG island near the promoter is fully methylated. In cell types in which the gene is expressed, the CpG island near the promoter is unmethylated. As a final example, the inactive X chromosome (Barr body) in female mammals contains methylated CpG islands adjacent to most structural genes; this high degree of CpG island methylation renders the Barr body transcriptionally silent.

### Key Questions

- How does methylation alter the structure of DNA?
- Where are many CpG islands located?
- In terms of DNA methylation, what is the difference between a housekeeping gene and a tissue-specific gene?

## Methylation Blocks Activator Proteins and Recruits HDACs

DNA methylation is thought to silence the transcription of a nearby gene in two general ways. First, methylation at a CpG island near the promoter of a gene prevents an activator protein from binding to an enhancer DNA sequence (see **figure 14.11**). DNA methylation inhibits activator binding because the methyl group on cytosine prevents the recognition helix (see above) within activator proteins from binding to the DNA major groove. Second, methylated CpG islands near promoters serve as the binding sites for **methyl-CpG-binding proteins**. When a methyl-CpG-binding protein binds to a methylated CpG island, the methyl-CpG-binding proteins recruit histone deacetylases (HDACs). HDACs then remove the acetyl groups from histone tails, converting the core promoter region of the gene into the heterochromatin (closed) state. Transcription of the nearby gene is therefore inhibited.

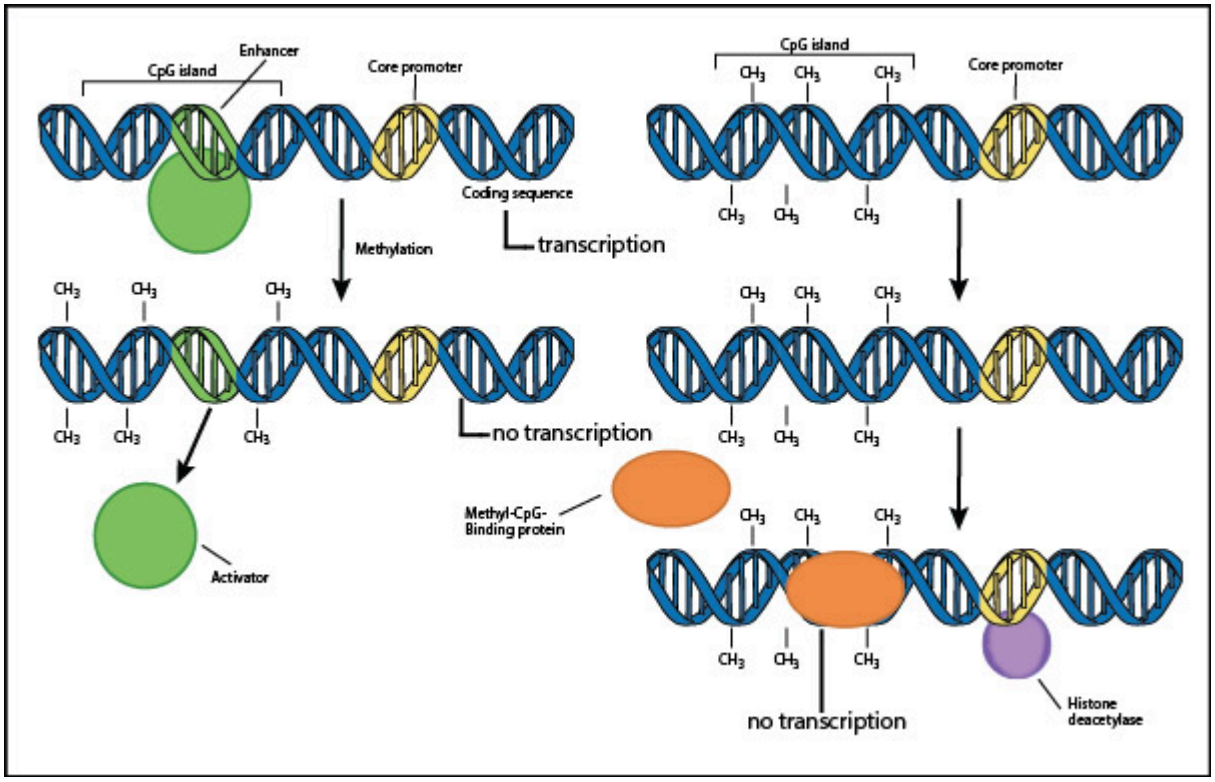



*Figure 14.11* ***Methylation Inhibits Transcription*** *--- Image created by SL*

## Key Questions

- Describe the two ways that DNA methylation can inhibit transcription.

# DNA Methylation is Preserved During Cell Division

The DNA methylation pattern in the cell is established by a process called ***de novo* methylation** (see **figure 14.12**). *De novo* methylation converts unmethylated DNA to fully methylated DNA (i.e., both DNA strands are methylated). *De novo* methylation is thought to occur during embryonic development or when cells differentiate to form tissues. Unfortunately, the details of *de novo* methylation are poorly understood.

The DNA methylation pattern established during *de novo* methylation is preserved during cell division; if a CpG island is fully methylated in a cell prior to mitosis, the same CpG island is fully methylated in the two daughter cells at the conclusion of mitosis. **Maintenance methylation** ensures that the daughter cells produced by mitosis maintain the same methylation pattern as the parental cell. For instance, suppose that fully methylated DNA is replicated. Because the DNA replication machinery does not methylate nitrogenous bases, the daughter DNA strands produced do not contain methylated cytosine bases. Thus, the daughter double-stranded DNA molecules are initially **hemimethylated**, with a methylated parental strand and an unmethylated daughter DNA strand. This hemimethylated DNA is recognized by **DNA methyltransferase**, which subsequently methylates the cytosine bases on the daughter DNA strands, thus preserving the DNA methylation pattern established in the parental cell.

Methylation of DNA explains a genetic phenomenon called **genomic imprinting**. In oogenesis (egg cell formation) or spermatogenesis (sperm cell formation), a specific gene is methylated by *de novo* methylation. Following fertilization, the methylation pattern is maintained as the fertilized egg begins to divide. For example, if the paternal allele for a gene

is fully methylated by genomic imprinting, that paternal allele remains fully methylated in the cells of the offspring. We will discuss genomic imprinting more in Part 15.

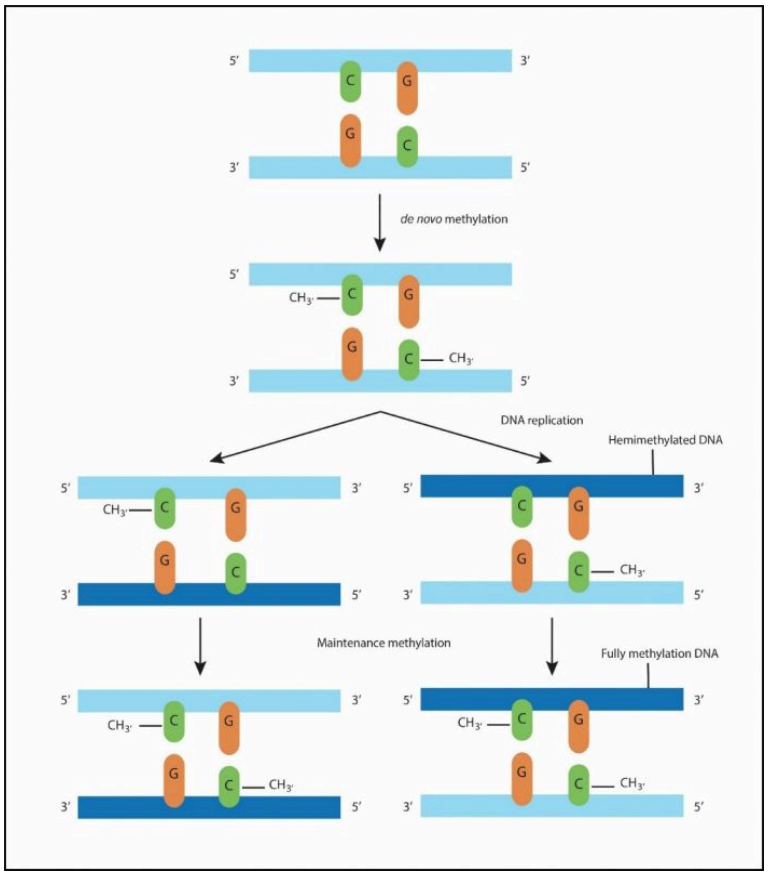



*Figure 14.12* ***Preserving DNA Methylation During Cell Division*** *--- image created by SL*

## Key Questions

- What is the difference between *de novo* and maintenance methylation?
- Which enzyme is responsible for maintenance methylation?
- What is meant by genomic imprinting?

# Insulators

In eukaryotes, the processes that regulate the expression of one structural gene, such as activators/repressor proteins binding, histone acetylation, and DNA methylation do not necessarily influence the regulation of an adjacent gene. **Insulator** DNA sequences define the boundaries between genes (see **figure 14.13**); an insulator DNA sequence ensures that the gene regulation processes that affect one gene do not affect nearby genes. Insulator DNA sequences:

- **Serve as the binding sites for proteins that act as physical barriers for the HATs, HDACs and SWI/SNF protein complexes.** For example, suppose a gene is flanked by two insulator DNA sequences, and HATs modify histone tails and activate transcription of the gene. Because the proteins bound to insulators serve as physical barriers to the HATs, genes beyond the insulator sequences are not activated.
- **Serve as the binding sites for proteins that limit the effects of enhancer/silencer sequences.** Suppose that Gene A has an adjacent enhancer DNA sequence. Gene B is also near the enhancer DNA sequence. A protein bound to the insulator DNA sequence between Genes A and B ensures that the enhancer only activates Gene A; the transcription of Gene B is unaffected. Insulators can limit the effects of silencer DNA sequences in a similar way.

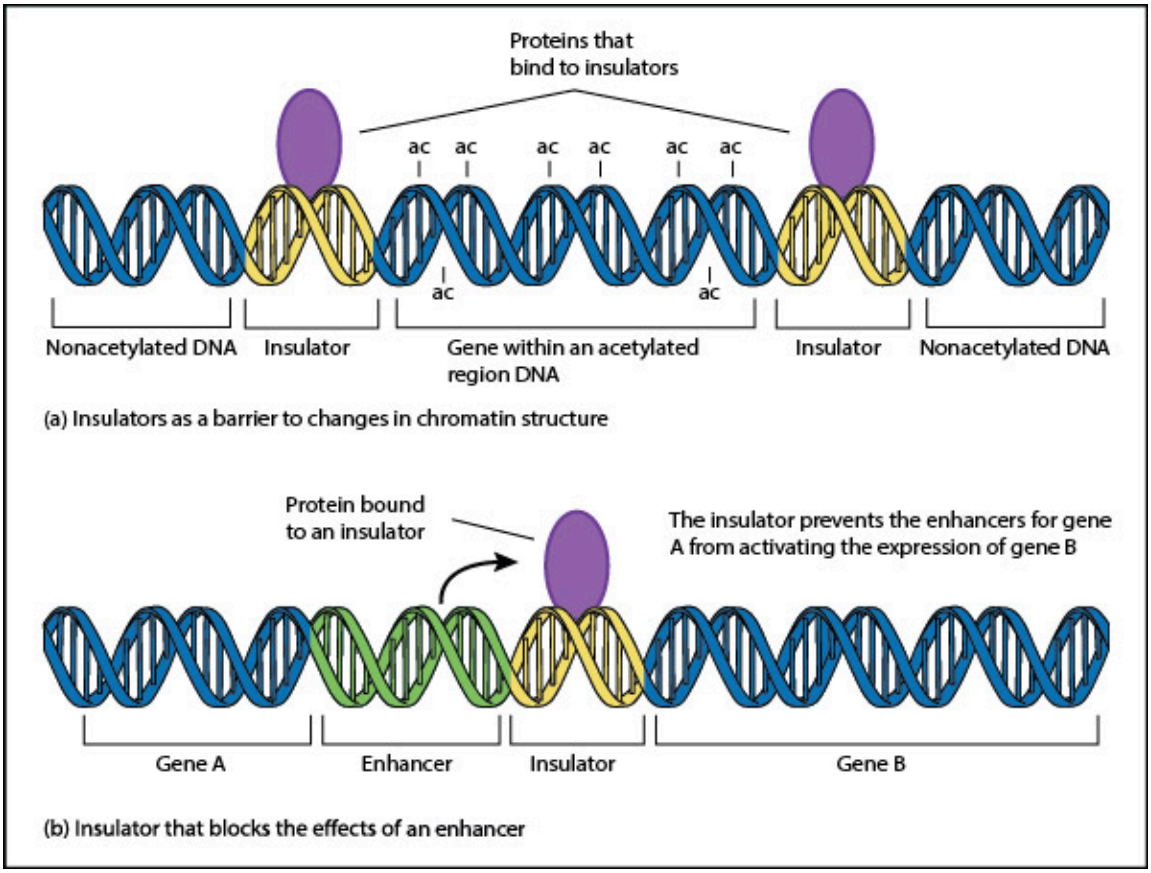



*Figure 14.13* ***Insulators*** *--- Image created by SL*

## Key Questions

- What is an insulator DNA sequence?
- How do insulators ensure that gene regulation is limited to a single gene?

# Part 14 Review

**Fill in the blank:**

1. The core promoter consists of two consensus DNA sequences located at position _____ and _____.
2. The general transcription factor (GTF) proteins are _____.
3. Some examples of regulatory transcription factor proteins are _____, which increase transcription and _____, which decrease transcription below basal levels.
4. Transcription factor proteins contain structural motifs. Two transcription factors with the _____ motif interact and form a coiled coil. Two alpha-helices are part of a _____ motif seen in proteins involved in muscle cell differentiation.
5. The interaction of two identical transcription factor proteins to produce one molecule is called a _____.
6. One example of a steroid hormone is _____.
7. A glucocorticoid receptor is bound to _____ until a glucocorticoid hormone molecule binds to the receptor.
8. CREB is a (protein OR DNA sequence; circle the correct answer), whereas CRE is a (protein OR DNA sequence; circle the correct answer).
9. Upon its activation in the CREB system, protein kinase A (PKA) enters the nucleus and phosphorylates CREB which then leads to transcription being (turned ON or turned OFF; circle the correct answer).
10. Histone acetyltransferases add acetyl groups to _____ amino acids on the histone tail.
11. Acetyl groups are removed from histone tails by enzymes called _____.
12. Methyl groups added to cytosine bases usually project into the (major OR minor; circle the correct answer) groove of the DNA.
13. Housekeeping genes are usually (methylated OR unmethylated; circle the correct answer) while tissue-specific genes are (methylated or unmethylated; circle the correct answer) in cells that do not express the gene.
14. _____ methylation ensures that the methylation pattern continues in the daughter cells produced by mitosis.