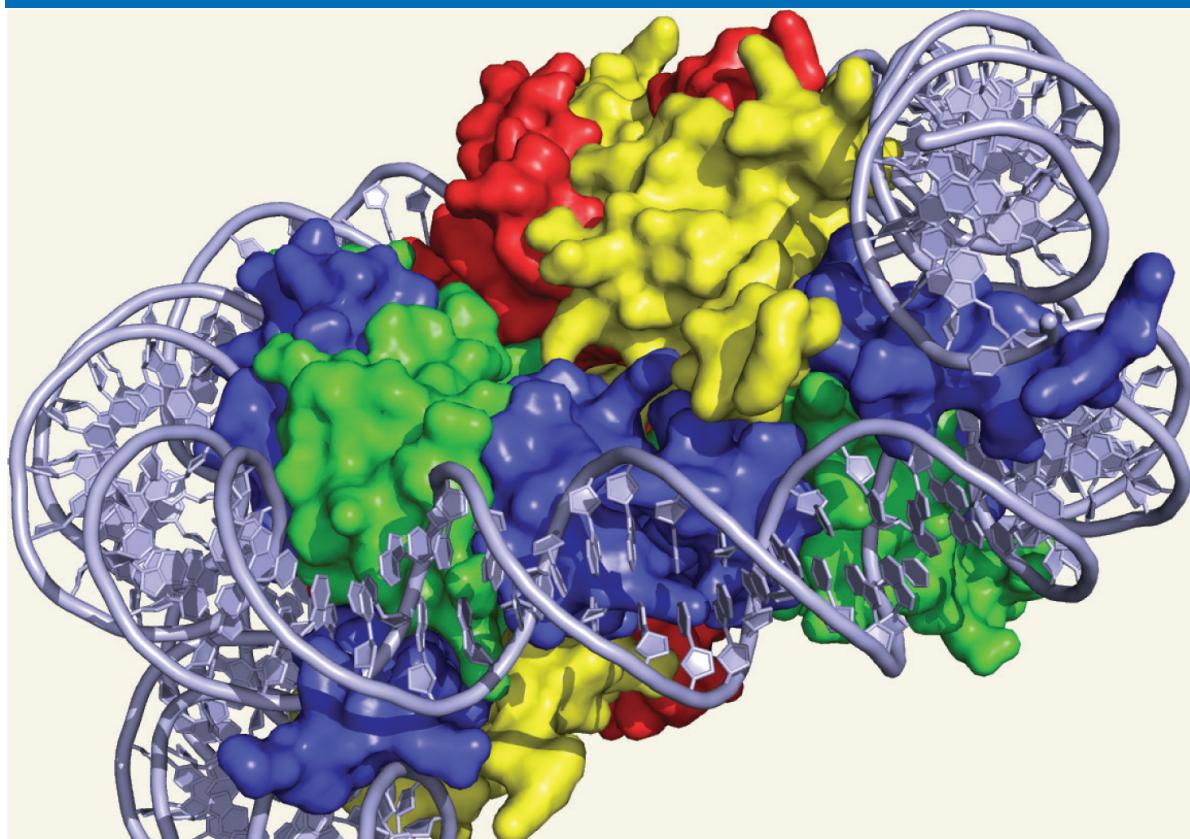


CHAPTER 12

Regulation of Transcription in Eukaryotes



PDB ID 5y0c.

In eukaryotic cells, DNA (gray) is wrapped around histone proteins (blue, green, red, and yellow), which affects access of the transcription machinery to DNA.

CHAPTER OUTLINE AND LEARNING OBJECTIVES

12.1 TRANSCRIPTION FACTORS REGULATE TRANSCRIPTION

LO 12.1 Diagram how transcription factors and DNA enhancer elements control the transcription of individual genes.

12.2 CHROMATIN STRUCTURE

LO 12.2 Draw a segment of chromatin, labeling each histone, a nucleosome, and the structural features that are important to their function in transcription.

12.3 CHROMATIN REGULATES TRANSCRIPTION

LO 12.3 Compare and contrast how chromatin modifying and chromatin remodeling mechanisms contribute to gene-specific transcription regulation.

12.4 CHROMATIN IN EPIGENETIC REGULATION

LO 12.4 Provide examples of the chromatin-based mechanisms that maintain gene expression over cellular or organismal generations.

CHAPTER OBJECTIVE

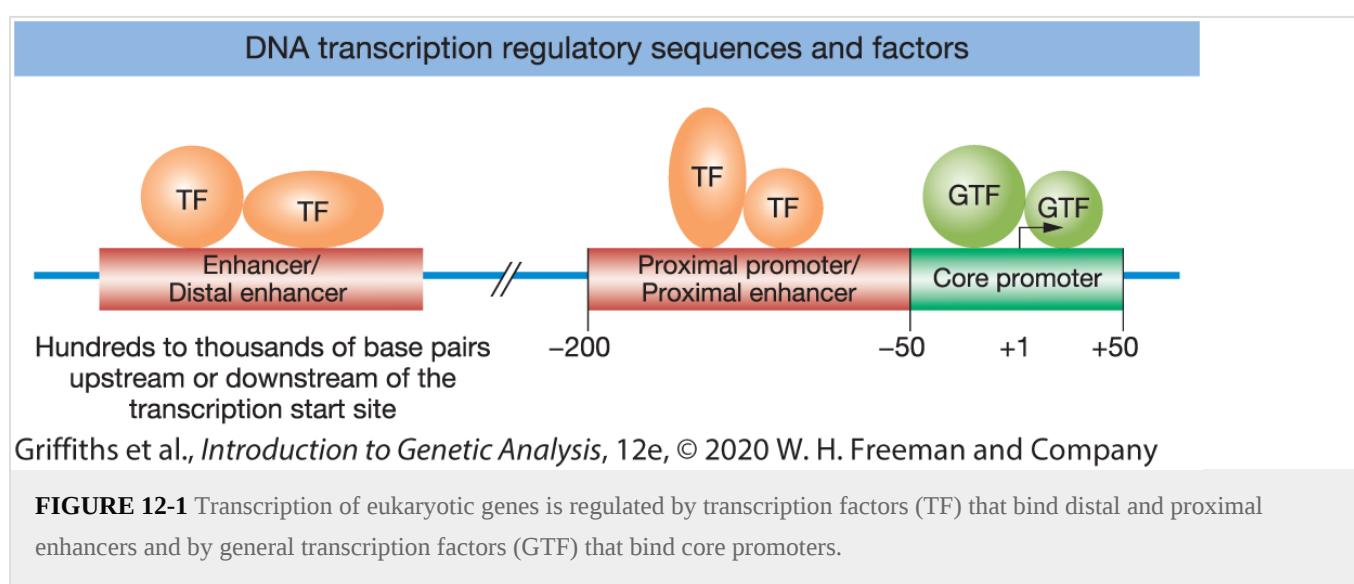
The process of transcription in bacterial and eukaryotic cells is carried out by fundamentally similar mechanisms, as described in [Chapter 8](#). In contrast, the regulation of transcription in eukaryotic cells is more complex than in bacterial cells. This increased complexity is mediated by a larger variety of regulatory DNA sequence elements and protein factors, including proteins that package DNA to fit into the nucleus. The main goal of this chapter is to describe how the variety of DNA sequence elements and protein factors in eukaryotic cells function in different combinations to precisely control the transcription of individual genes.

In [Chapters 8](#) and [11](#), you learned that transcription in bacteria is often regulated by single activator or repressor proteins that directly bind DNA. Initial expectations were that eukaryotic transcription would be regulated by similar means. However, in most eukaryotes, multiple proteins and DNA sequences work together to control transcription. A key additional difference between bacteria and eukaryotes is that in eukaryotes access to transcription regulatory sequences in DNA is restricted by the packaging of DNA with proteins in the nucleus. Gene regulation in eukaryotes involves proteins that promote or restrict access of RNA polymerases to gene promoters. This chapter will focus on the transcription of protein-coding genes by RNA polymerase II and thereby provide the foundation for understanding the regulation of transcription in time and space that choreographs the process of development described in [Chapter 13](#).

12.1 TRANSCRIPTION FACTORS REGULATE TRANSCRIPTION

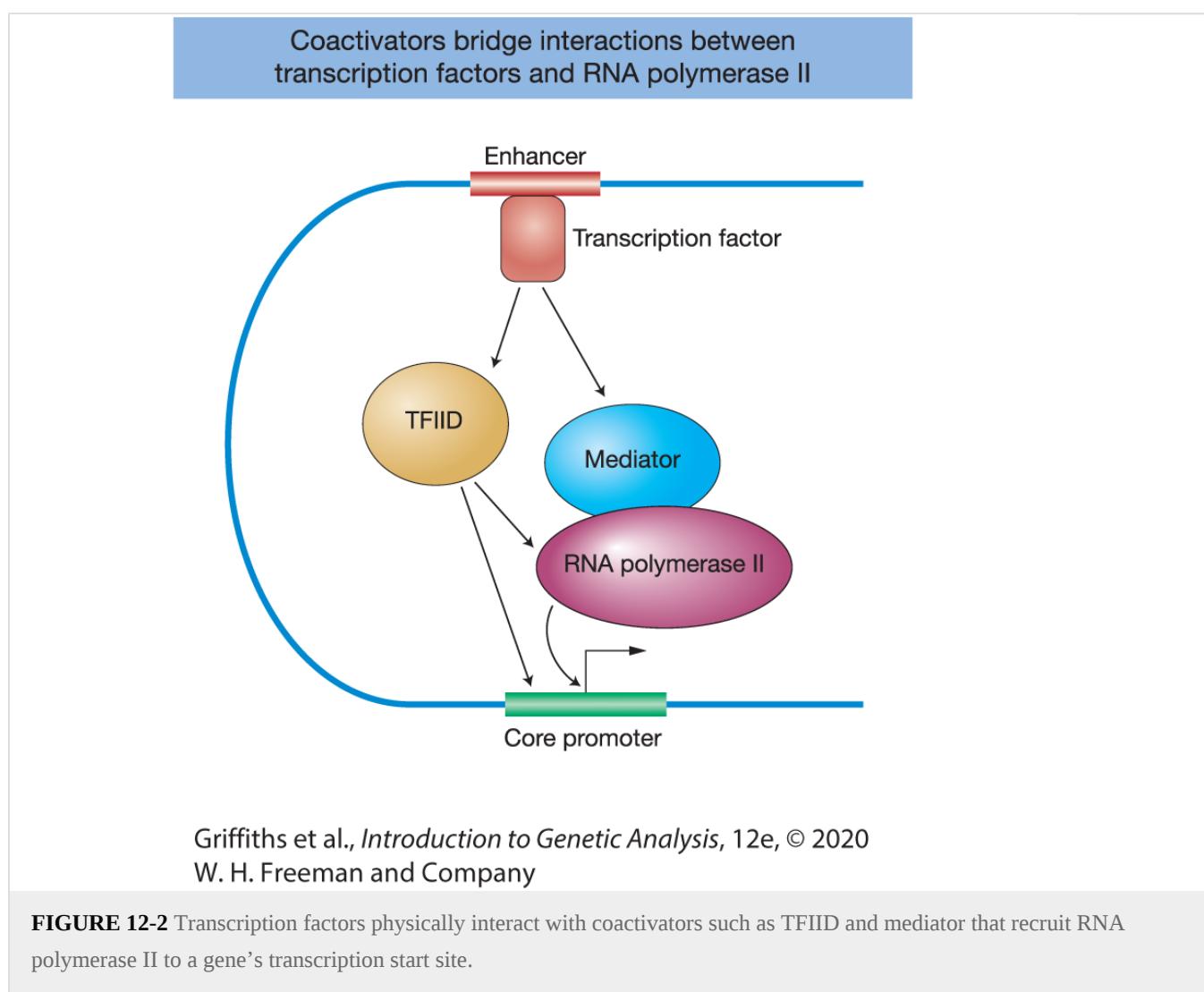
LO 12.1 Diagram how transcription factors and DNA enhancer elements control the transcription of individual genes.

The machinery required for generating the distinct patterns of gene transcription that occur in eukaryotic cells has many components, including trans-acting regulatory proteins and cis-acting regulatory DNA sequences. The regulatory proteins can be divided into two sets, those that directly bind DNA and those that do not. The first set of regulatory proteins consists of **transcription factors** that directly bind regulatory DNA sequences called **enhancers**. Enhancers that are located close to the core promoter are part of **proximal promoters** and are called **proximal enhancers**, and those that are a considerable distance from the promoter are part of **distal enhancers** and are called enhancers (**Figure 12-1**). In addition, some **general transcription factors (GTFs)** directly bind DNA regulatory sequences within **core promoters** that surround transcription start sites.



The second set of regulatory proteins consists of coregulators, which do not directly bind DNA. There are two types of coregulators: **coactivators** and **corepressors**. Coactivators and corepressors, respectively, increase or decrease the amount of transcription through binding or

enzymatically modifying other transcription regulatory factors. For example, some coactivators serve to bridge transcription factors and RNA polymerase II ([Figure 12-2](#)), while others alter the structure of chromatin, which is described later in this chapter.



KEY CONCEPT Distal and proximal enhancers are DNA sequences that regulate the transcription of genes. Coregulators, which bind transcription factors, control the recruitment and access to DNA of general transcription factors and RNA polymerase II.

Eukaryotic transcription regulatory mechanisms have been discovered through both biochemical and genetic approaches. The latter has been advanced in particular by studies of the single-celled yeast *Saccharomyces cerevisiae* (see the [Model Organism box below](#)). This organism, which has played a key role in wine making, beer making, and baking for many centuries, has been a

passport to understanding much of eukaryotic molecular biology. Several decades of research have produced many fundamental insights into general principles of how eukaryotic transcription regulatory proteins work.

MODEL ORGANISM

Yeast



SCIMAT/Science Source.

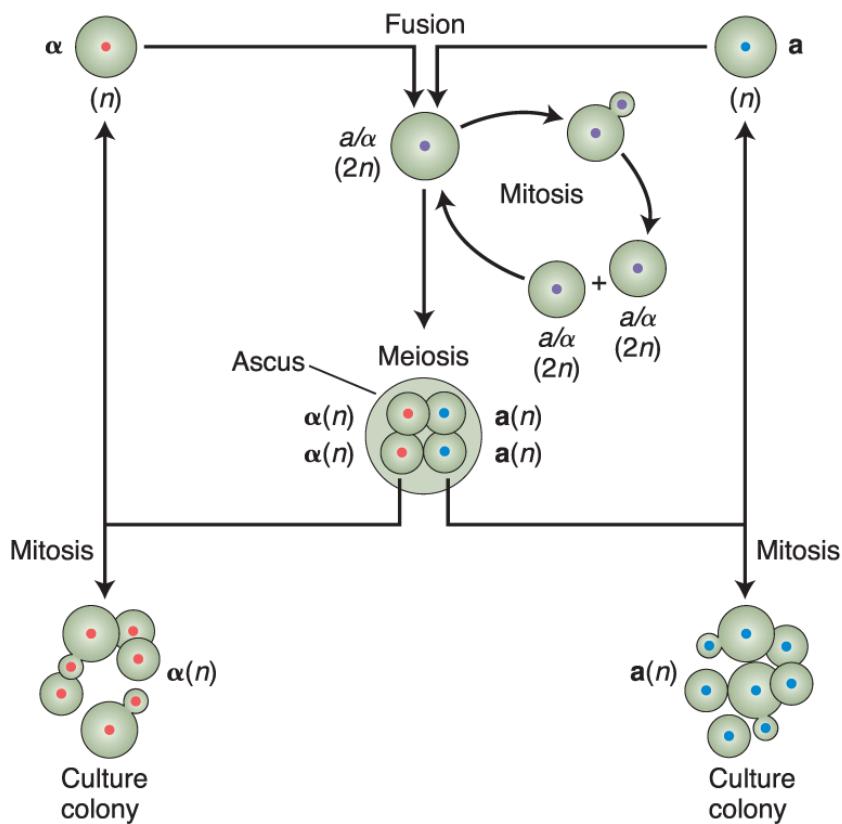
Electron micrograph of budding yeast cells.

Saccharomyces cerevisiae, or budding yeast, is a premier eukaryotic genetic system. Humans have grown yeast for centuries because it is an essential component of beer, bread, and wine. Yeast has many features that make it an ideal model organism. As a unicellular eukaryote, it can be grown on agar plates and, with yeast's life cycle of just 90 minutes, large quantities can be cultured in liquid media. It has a very compact genome with only about 12 megabase pairs of DNA (compared with almost 3000 megabase pairs for humans) containing approximately 6000 genes that are distributed on 16 chromosomes. It was the first eukaryote to have its genome sequenced.

The yeast life cycle makes it very versatile for laboratory studies. Cells can be grown as either diploids or haploids. In both cases, the mother cell produces a bud containing an identical daughter cell. Diploid cells either continue to grow by budding or are induced to undergo meiosis, which produces four haploid spores held together in an ascus (also called a tetrad). Haploid spores of opposite mating type (**a** or **α**) will fuse and form a diploid. Spores of the same mating type will continue growth by budding.

Yeast has been called the *E. coli* of eukaryotes because of the ease of forward and reverse mutant analysis. To isolate mutants using a forward genetic approach, haploid cells are mutagenized (with X rays, for example) and screened on plates for mutant phenotypes. This procedure is usually done by first plating cells on a rich medium on which all cells grow and by copying, or replica plating, the colonies from this master plate onto replica plates containing selective media or special growth conditions. For example, temperature-sensitive mutants will grow on the master plate at the permissive temperature but not on a replica plate at a restrictive temperature. Comparison of the colonies on master and replica plates will reveal the temperature-sensitive mutants. Using

reverse genetics, scientists can replace any yeast gene of known or unknown function with a mutant version to understand the nature of the gene product.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

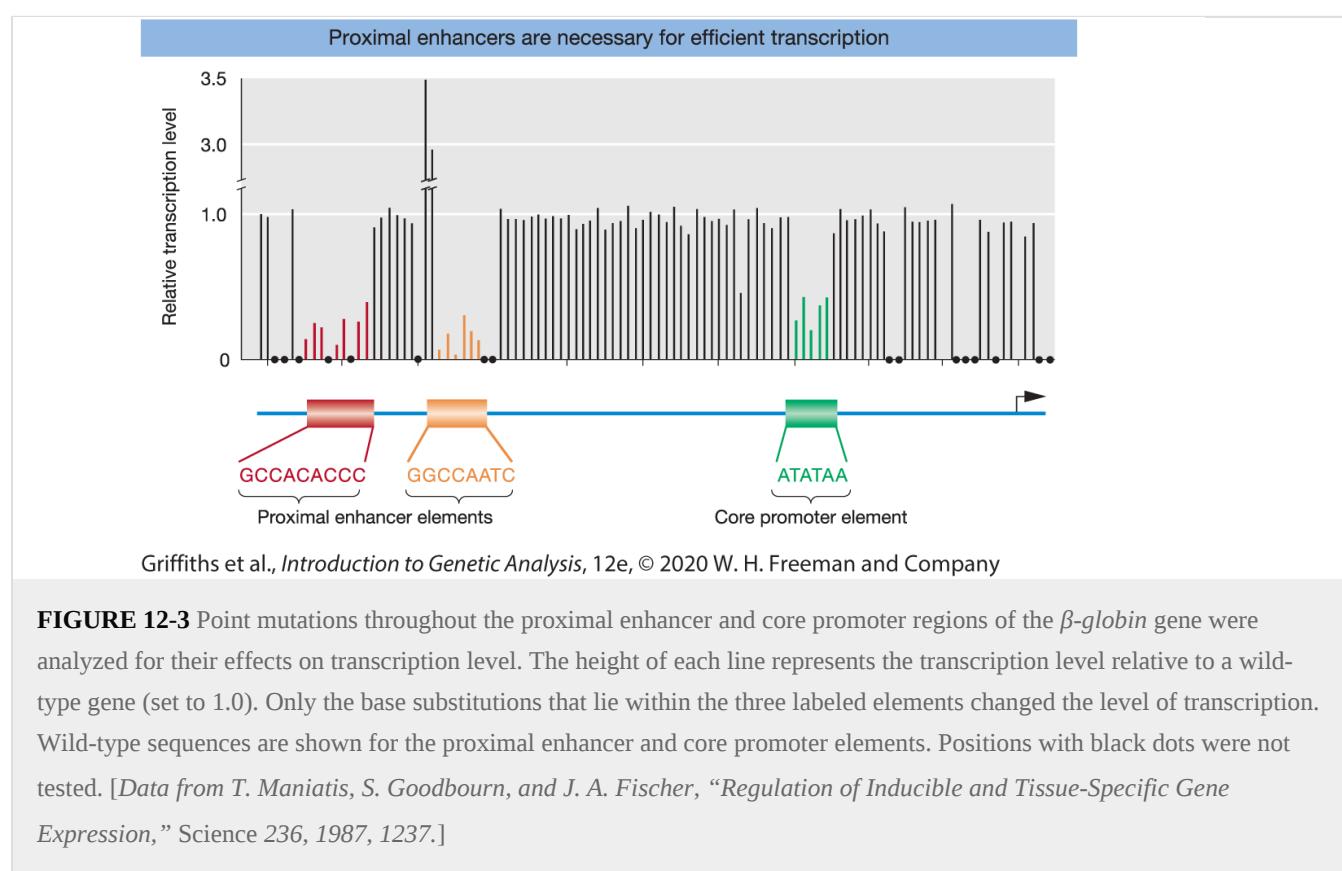
The life cycle of baker's yeast. The nuclear alleles *MATa* and *MATα* determine mating type. Ploidies, n and $2n$, are indicated in parentheses.

Fu

Transcription factors bind distal and proximal enhancers

Mutagenesis studies have revealed the importance of proximal enhancers. As shown in [Figure 12-3](#), point mutations in proximal enhancers as well as core promoters reduce transcription of the β -globin gene. This example reveals general features of enhancers: they contain short sequence elements (6–10 base pairs), and multiple elements are often clustered together. Enhancer elements frequently occur as inverted repeats of the same DNA sequence for binding of two similar or identical transcription factors, reminiscent of the DNA sequences controlling the *lac* operon in

bacteria ([Figure 11-14](#)). Because enhancer elements are short, they randomly occur many times in genomes. However, they are not all bound by transcription factors because binding often requires interactions with partner transcription factors bound to other nearby enhancers.

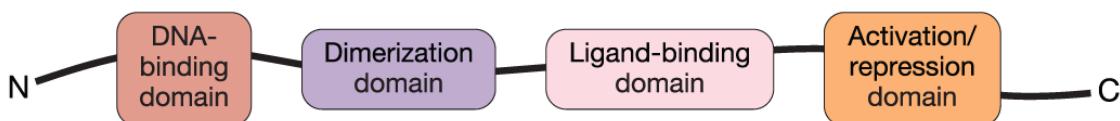


Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-3 Point mutations throughout the proximal enhancer and core promoter regions of the β -globin gene were analyzed for their effects on transcription level. The height of each line represents the transcription level relative to a wild-type gene (set to 1.0). Only the base substitutions that lie within the three labeled elements changed the level of transcription. Wild-type sequences are shown for the proximal enhancer and core promoter elements. Positions with black dots were not tested. [Data from T. Maniatis, S. Goodbourn, and J. A. Fischer, “Regulation of Inducible and Tissue-Specific Gene Expression,” *Science* 236, 1987, 1237.]

In addition to binding DNA enhancer elements, transcription factors bind other proteins ([Figure 12-4](#)). This is exemplified by C/EBP (CCAAT/enhancer-binding protein), the transcription factor that binds one of the proximal enhancer elements in the β -globin gene, the CCAAT box (pronounced “cat” box). C/EBP is characterized by a **DNA-binding domain** and a **dimerization domain**, the latter of which facilitates the formation of homodimers (binding between two C/EBP proteins) and heterodimers (binding between different C/EBP family members). C/EBP also contains an **activation domain** that interacts with other components of the transcription machinery to turn on transcription. Other transcription factors have **repression domains** that use similar mechanisms to turn off transcription. Furthermore, some transcription factors include a **ligand-binding domain** that binds a ligand such as a hormone or a vitamin, changing the structure of the transcription factor and activating it. As an example, binding of the hormone estrogen by a transcription factor called Estrogen receptor in the cytoplasm leads to its dimerization, nuclear localization, and binding to enhancers elements called Estrogen Response Elements. All transcription factors contain a DNA-binding domain and an activation/repression domain, but only some transcription factors contain dimerization and ligand-binding domains.

Transcription factors have multiple domains



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-4 Transcription factors can have four types of functional domains. All transcription factors contain a DNA-binding domain and an activation/repression domain. Some also contain a dimerization and/or a ligand-binding domain. Note that the order of the domains can differ in different transcription factors.

KEY CONCEPT Transcription factors use their DNA-binding, activation/repression, dimerization, and ligand-binding domains to activate or repress gene transcription.

Transcription factors: lessons from the yeast GAL system

Yeast make use of extracellular galactose (gal) by importing and converting it into a form of glucose that can be metabolized. Several genes—*GAL1*, *GAL2*, *GAL7*, and *GAL10*—in the yeast genome encode enzymes that catalyze steps in this metabolic pathway ([Figure 12-5](#)). Three additional genes—*GAL3*, *GAL4*, and *GAL80*—encode proteins that regulate transcription of the enzyme-encoding genes. Just as in the *lac* system of *E. coli*, the abundance of the sugar determines the level of transcription in the metabolic pathway. In yeast cells growing in media lacking galactose, the *GAL* genes are largely transcriptionally silent. But, in the presence of galactose (and the absence of glucose), the *GAL* genes are transcriptionally induced. Just as for the *lac* operon, genetic and molecular analyses of mutants have been key to understanding how transcription of genes in the galactose pathway is controlled.

The Gal pathway

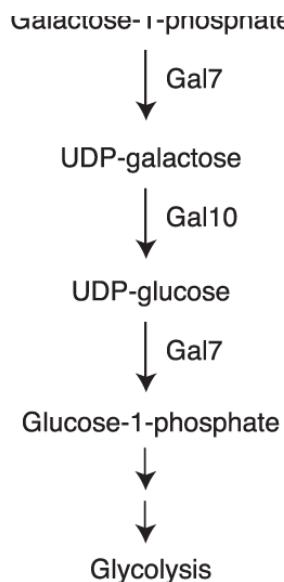
Galactose (extracellular)

↓
Gal2

Galactose (intracellular)

↓
Gal1

Galactose → Glucose → Energy



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 12-5 Galactose is converted into glucose-1-phosphate in a series of biochemical steps. These steps are catalyzed by the enzymes Gal1, Gal2, Gal7, and Gal10, which are encoded by the genes *GAL1*, *GAL2*, *GAL7*, and *GAL10*, respectively.

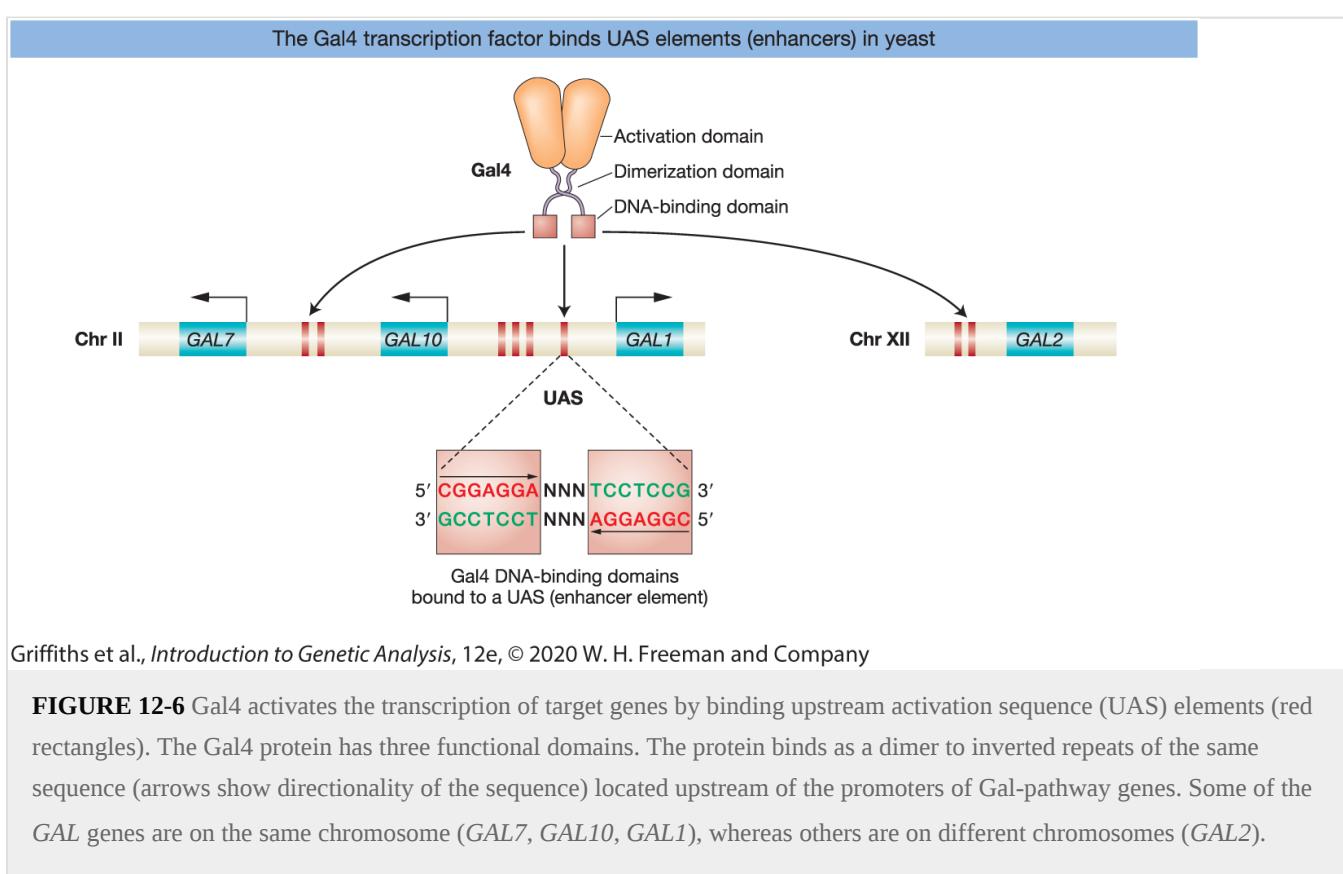
ANIMATED ART Sapling Plus

Galactose metabolism in yeast

Gal4 binds enhancers called upstream activation sequences

The key regulator of *GAL* gene transcription is the Gal4 transcription factor, a sequence-specific DNA-binding protein. In the presence of galactose, mRNA levels for the *GAL1*, *GAL2*, *GAL7*, and *GAL10* genes are about 1000-fold higher than in its absence. However, in *GAL4* mutants they are unchanged, indicating that Gal4 is required for transcription of these genes. Each of the four genes has two or more Gal4-binding sites (i.e., enhancers) located at some distance 5' (upstream) of its promoter ([Figure 12-6](#)). Consider the *GAL10* and *GAL1* genes, which are adjacent to each other and transcribed in opposite directions. Between the *GAL1* and *GAL10* transcription start sites is a 118-base-pair region that contains four Gal4-binding sites. Each Gal4-binding site is 17 base pairs long and is bound by a homodimer of Gal4 proteins (two Gal4 proteins bound together). There are also two Gal4-binding sites upstream of the *GAL2* gene and another two upstream of the *GAL7* gene. These binding sites are required for transcription activation. If the binding sites are deleted,

the genes are transcriptionally silent, even in the presence of galactose. Because the Gal4 enhancers are located upstream of the genes they regulate, they are also called **upstream activation sequence (UAS)** elements.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-6 Gal4 activates the transcription of target genes by binding upstream activation sequence (UAS) elements (red rectangles). The Gal4 protein has three functional domains. The protein binds as a dimer to inverted repeats of the same sequence (arrows show directionality of the sequence) located upstream of the promoters of Gal-pathway genes. Some of the *GAL* genes are on the same chromosome (*GAL7*, *GAL10*, *GAL1*), whereas others are on different chromosomes (*GAL2*).

ANIMATED ART Sapling Plus

Gal4 regulation of galactose-metabolizing enzymes

KEY CONCEPT Transcription factors coordinately regulate the transcription of multiple genes involved in the same biological process by binding enhancers that are common to the genes.

In addition to its action in yeast cells, Gal4 has been shown to activate transcription of UAS-containing genes when they are introduced into insect cells, human cells, and many other eukaryotic organisms. This versatility suggests that transcription machineries and mechanisms of gene activation are common to a broad array of eukaryotes, and that features revealed in yeast are generally present in other eukaryotes and vice versa. Furthermore, because of their versatility, Gal4 and its UAS elements have become favored tools for manipulating gene expression in a wide variety of model organisms.

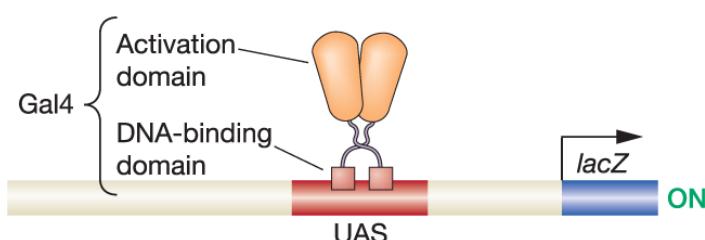
KEY CONCEPT The ability of Gal4 to function in a variety of eukaryotes indicates that eukaryotes generally have common transcription regulatory machineries and mechanisms.

Gal4 domains function independently of one another

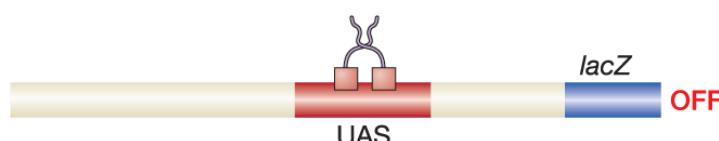
In addition to the DNA-binding domain and dimerization domain, Gal4 has an activation domain. A series of simple, elegant experiments demonstrated that the DNA-binding and activation domains of Gal4 as well as other transcription factors are modular; that is, they function independently of one another ([Figure 12-7](#)). In this study, researchers fused the Gal4 activation domains to the DNA-binding domain from the *E. coli* transcription factor LexA. Transcription was measured using reporter genes (see [Chapter 10](#)) that contained Gal4-binding sites (i.e., UAS) or LexA-binding sites (i.e., LexA site) upstream of a promoter and the *E. coli lacZ* gene coding region. The level of transcription of *lacZ* in yeast cells was determined by measuring the level of its encoded protein product β-galactosidase. Full-length Gal4 activated transcription when bound to the UAS ([Figure 12-7a](#)) but the Gal4 DNA-binding domain lacking the activation domain did not ([Figure 12-7b](#)). Similarly, the LexA DNA-binding domain did not activate transcription from LexA sites ([Figure 12-7c](#)), but a protein fusion of the Gal4 activation domain and the LexA DNA-binding domain did ([Figure 12-7d](#)). Likewise, a fusion of the Gal4 binding domains to another activation domain was able to activate transcription (not shown).

Transcription factors are modular in structure

(a) The complete Gal4 dimer

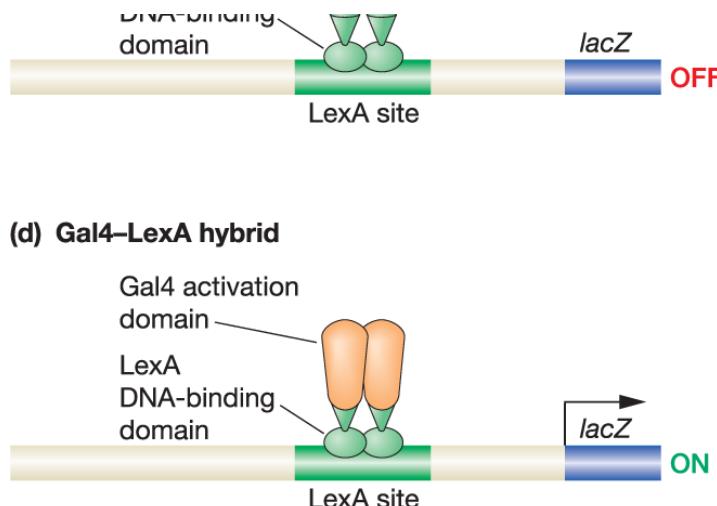


(b) Gal4 lacking the activation domain



(c) LexA lacking the activation domain

DNA-binding



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 12-7 Transcription factors have multiple, separable domains. (a) Full-length Gal4 has three domains and activates transcription from UAS sites. (b) Removal of the Gal4 activation domain shows that dimerization and DNA binding is not sufficient for transcription activation. (c) Similarly, the LexA DNA-binding domain cannot activate transcription, but (d) when fused to the Gal4 activation domain, it can activate transcription through LexA-binding sites.

ANIMATED ART Sapling Plus

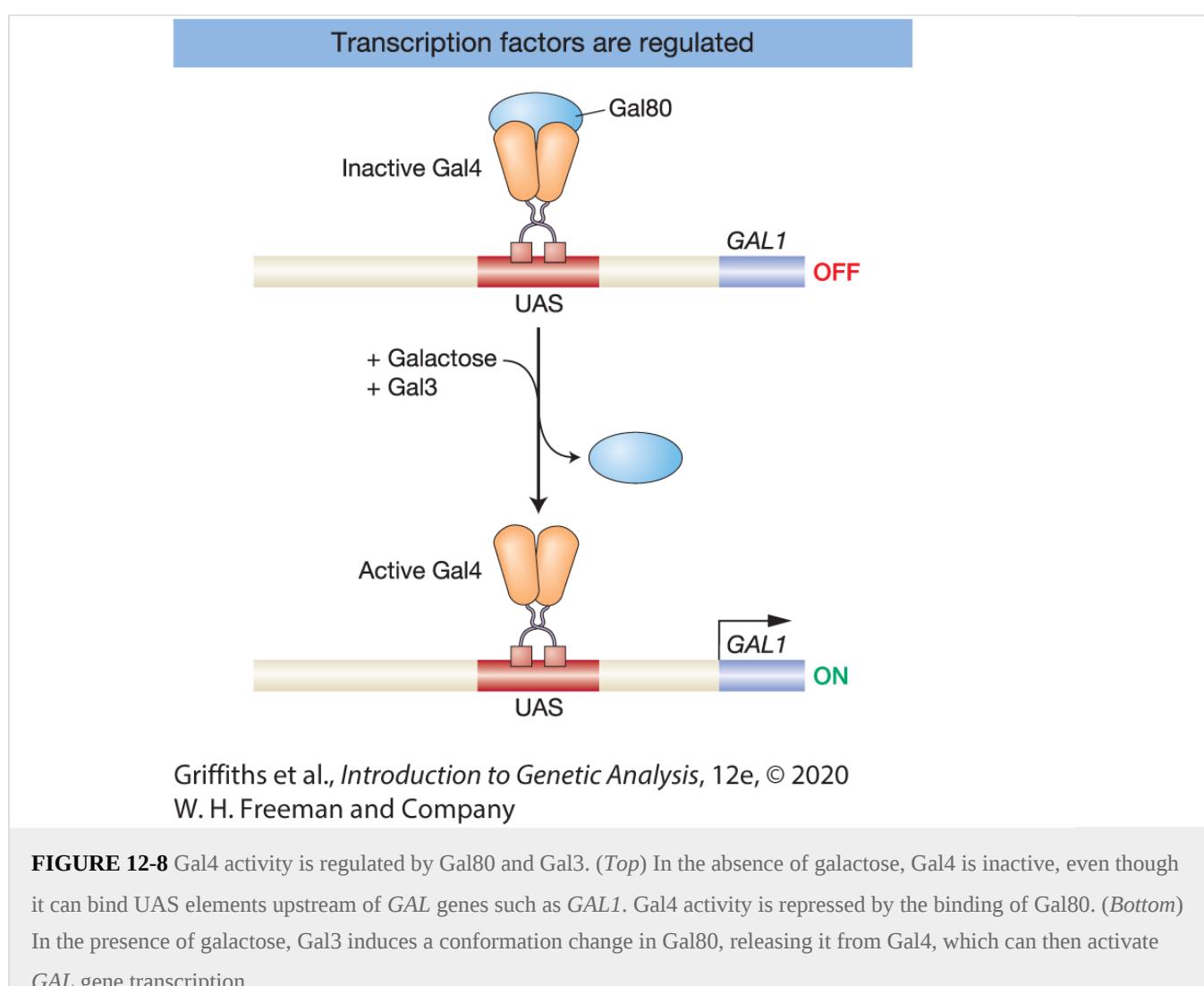
Gal4 modularity and gene induction in non-yeast

Researchers have used the modularity of transcription factors to develop technologies such as the yeast two-hybrid system that is used to detect protein-protein interactions *in vivo* ([Chapter 14](#)). The modularity of transcription factors is also the cause of some cancers such as acute promyelocytic leukemia (APL), a cancer of early blood-forming cells. In almost all cases of APL, a chromosome translocation creates a gene fusion between the activation domain of PML (promyelocytic leukemia) and the DNA-binding and ligand binding domains of RARA (retinoic acid receptor α). The fusion protein assembles with corepressor proteins, instead of coactivator proteins, to block transcription of normal RARA gene targets that control the differentiation of myeloid (blood) cells, which leads to uncontrolled proliferation of these cells.

KEY CONCEPT Eukaryotic transcription factors are modular, having separable domains for DNA binding, activation/repression, dimerization, and ligand-binding.

Regulation of Gal4

Gal4 activity is regulated by the Gal80 and Gal3 proteins ([Figure 12-8](#)). In *GAL80* mutants, the *GAL* structural genes (*GAL1*, *GAL2*, *GAL7*, and *GAL10*) are transcriptionally active even in the absence of galactose. This suggests that the normal function of Gal80 is to inhibit *GAL* gene transcription. Conversely, in *GAL3* mutants, the *GAL* structural genes are not active in the presence of galactose, suggesting that Gal3 normally promotes transcription of the *GAL* genes.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 12-8 Gal4 activity is regulated by Gal80 and Gal3. (Top) In the absence of galactose, Gal4 is inactive, even though it can bind UAS elements upstream of *GAL* genes such as *GAL1*. Gal4 activity is repressed by the binding of Gal80. (Bottom) In the presence of galactose, Gal3 induces a conformation change in Gal80, releasing it from Gal4, which can then activate *GAL* gene transcription.

ANIMATED ART  Sapling Plus

Gal4 regulation of galactose-metabolizing enzymes

Extensive biochemical analyses revealed that Gal80 is a corepressor of Gal4. Gal80 binds Gal4 with high affinity and directly inhibits Gal4 activity. Specifically, Gal80 binds within the Gal4 activation domain, blocking its ability to promote transcription. Gal80 is expressed continuously, so it is always acting to repress transcription of the *GAL* structural genes unless stopped.

The role of Gal3 is to release the *GAL* structural genes from their repression by Gal80 when galactose is present. Gal3 is thus both a sensor and inducer. When Gal3 binds galactose and ATP, it undergoes a conformational change that promotes binding to Gal80, which in turn causes Gal80 to be released from Gal4, which is then able to interact with coactivators and RNA polymerase II to activate transcription. Thus, Gal80, Gal3, and Gal4 are all part of a switch, whose state is determined by the presence or absence of galactose ([Figure 12-8](#)). In this switch, DNA binding by the transcription factor is not the physiologically regulated step (as is the case in the *lac* operon and bacteriophage λ); rather, the ability of the activation domain to perform its function is regulated.

KEY CONCEPT Environmental signals such as galactose alter the activity of eukaryotic transcription factors by controlling their interactions with other proteins.

Combinatorial control of transcription: lessons from yeast mating type

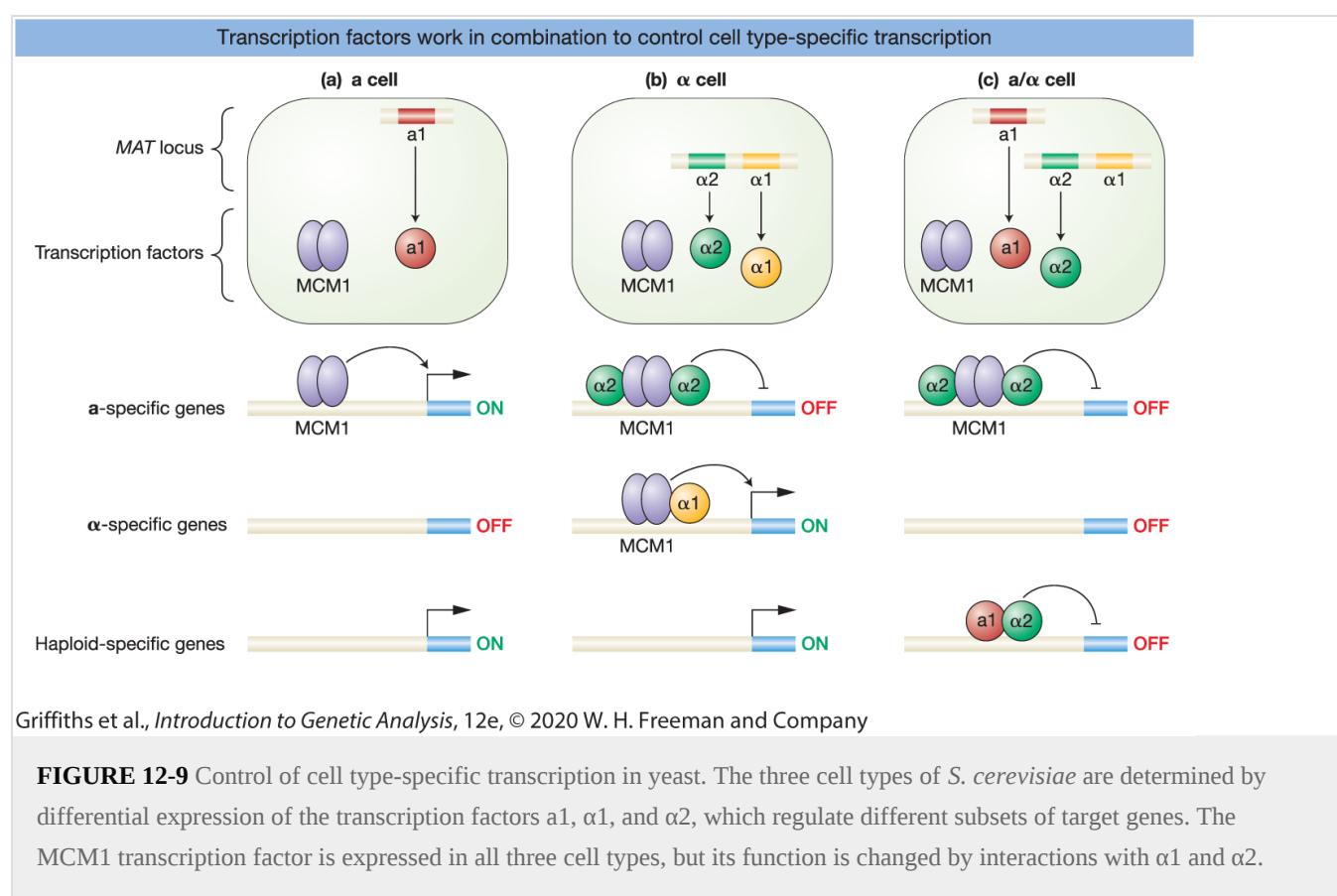
Thus far, we have focused on transcription regulation of single genes or a few genes in one pathway. In multicellular organisms, distinct cell types differ in the transcription of hundreds of genes. The activation or repression of sets of genes must therefore be coordinated in the making of particular cell types. One of the best-understood examples of cell type-specific regulation in eukaryotes is the regulation of mating type in yeast. This regulatory system has been dissected by a combination of genetics, molecular biology, and biochemistry. Mating type serves as an excellent model for understanding the logic of transcription regulation in multicellular animals.

The yeast *Saccharomyces cerevisiae* can exist in any of three different cell types known as **a**, **α** (alpha), and **a/α** . The two cell types **a** and **α** are haploid and contain only one copy of each chromosome. The **a/α** cell is diploid and contains two copies of each chromosome. Although the two haploid cell types cannot be distinguished by their appearance, they can be differentiated by a number of specific cellular characteristics, principally their mating type (see the [Model Organism](#)

[box above](#)). An **a** cell mates only with an **a** cell, and an **a** cell mates only with an **a** cell. An **a** cell secretes an oligopeptide pheromone, or sex hormone, called **a** factor that arrests **a** cells in the cell cycle. Similarly, an **a** cell secretes a pheromone, called **a** factor, that arrests **a** cells. Cell arrest of both participants is necessary for successful mating. The diploid **a/a** cell does not mate, is larger than **a** and **a** cells, and does not respond to mating hormones.

Genetic analysis of mutants defective in mating has shown that cell type is controlled by a single genetic locus, the mating-type locus, *MAT*. There are two alleles of the *MAT* locus: haploid **a** cells have the *MAT_a* allele, and haploid **a** cells have the *MAT_a* allele. The **a/a** diploid has both alleles. Yeast can switch mating type by a homologous recombination event that replaces one *MAT* allele with the other *MAT* allele. These two alleles activate different sets of genes because they encode different transcription factors. In addition, a transcription factor not encoded by the *MAT* locus, called MCM1, plays a key role in regulating cell type.

The simplest case is the **a** cell type ([Figure 12-9a](#)). The *MAT_a* locus encodes a single transcription factor, **a1**. However, **a1** has no effect in haploid cells, only in diploid cells. In a haploid **a** cell, the transcription factor MCM1 turns on the expression of the structural genes needed by an **a** cell by binding enhancers for **a**-specific genes.



In an α cell, the α -specific structural genes must be transcribed, but, in addition, MCM1 must be prevented from activating the a -specific genes ([Figure 12-9b](#)). The DNA sequence of the $MAT\alpha$ allele encodes two transcription factors, $\alpha 1$ and $\alpha 2$, that are produced by separate genes. These two proteins have different regulatory roles in the cell. The $\alpha 1$ protein is an activator of α -specific transcription. It binds in concert with the MCM1 protein to an enhancer that controls α -specific genes. The $\alpha 2$ protein represses transcription of the a -specific genes by binding with MCM1 to enhancers upstream of a -specific genes.

In a diploid yeast cell, transcription factors encoded by each MAT locus are expressed ([Figure 12-9c](#)). This results in repression of all genes involved in cell mating and a separate set of genes, called haploid specific, that are expressed in haploid cells but not diploid cells. The $a 1$ transcription factor encoded by $MATa$ has a part to play at last. $a 1$ can bind $\alpha 2$ and alter its binding specificity such that the $a 1-\alpha 2$ complex binds enhancers found upstream of haploid-specific genes and silences these genes. In diploid cells, then, the $\alpha 2$ protein exists in two forms: (1) as an $\alpha 2$ -MCM1 complex that represses a -specific genes, and (2) in a complex with the $a 1$ protein that represses haploid-specific genes. Moreover, the $a 1-\alpha 2$ complex also represses expression of the $\alpha 1$ gene, which is thus no longer present to turn on α -specific genes. Thus, cell type-specific transcription of genes that control mating type in yeast is achieved by multiple transcription factors working in different combinations.

KEY CONCEPT The control of yeast mating type is an example of how cell type-specific patterns of transcription in eukaryotes can be governed by different combinations of interacting transcription factors.

12.2 CHROMATIN STRUCTURE

LO 12.2 Draw a segment of chromatin, labeling each histone, a nucleosome, and the structural features that are important to their function in transcription.

In eukaryotic cells, DNA is packaged with proteins to create **chromatin**. In the cell's nucleus, DNA in chromatin is compacted over 10,000-fold compared to its linear form. The structure of chromatin serves to fit DNA into the nucleus, and it also serves as a substrate for reversible changes in protein-DNA and protein-protein interactions that regulate transcription. In this section, we describe the structure of **histones** (the major protein components of chromatin), **nucleosomes** (the basic structural units of chromatin), and higher-order chromatin structures (three-dimensional assemblies of nucleosomes with one another). Because higher-order chromatin structures in eukaryotic cells can make DNA inaccessible to binding by transcription factors, an understanding of chromatin structure is essential for understanding how transcription is regulated.

Histones

Eukaryotic cells express five types of histone proteins: H1, H2A, H2B, H3, and H4. Histones H2A, H2B, H3, and H4 are known as **core histones** because they form a core complex around which DNA is wrapped to form nucleosomes. Histone H1 is known as a **linker histone** because it binds the DNA that links adjacent nucleosomes. In addition to these **canonical histones** that package the newly replicated genome, there are **variant histones** that are incorporated into nucleosomes in a DNA-replication independent manner. As an example, the histone variant H2A-Z is 60 percent identical in sequence to canonical histone H2A. H2A-Z replaces H2A in nucleosomes at promoters of both transcriptionally active and silent genes. In contrast, another H2A variant called H2A-X is incorporated into nucleosomes at sites of DNA damage ([Chapter 15](#)). Variants of H1, H2B, and H3 also play specialized roles, but, as yet, no variants of H4 have been identified.

Histone proteins have unusual features that are relevant to their roles as structural and regulatory components of chromatin. They are extremely abundant. In mammalian cells, histones constitute approximately 70 percent of the protein complement of chromatin. There are about 10 million copies of each core histone per cell and about half this amount of histone H1. Core histone

proteins are small (11–15 kDa), unusually basic (at least 20 percent of their amino acids are lysine or arginine), and positively charged at neutral pH. Electrostatic interactions between positively charged amino acids and the negatively charged phosphate backbone of DNA play an important role in determining the structure of chromatin. The sequences of core histone proteins are among the most highly conserved in evolution. From yeast to humans, both H2A and H2B sequences are more than 70 percent identical, and both H3 and H4 are more than 90 percent identical. Because of this conservation, studies of histones in genetically controllable organisms such as yeast and *Drosophila* have provided considerable insights into the function of histones in higher eukaryotic organisms such as humans.

Core histone proteins have three types of structural domains: **histone folds**, **histone-fold extensions**, and **flexible tails** (Figure 12-10a). Histone folds are located in the central region of histone proteins. They are approximately 70 amino acids in length and are made up of three α -helices separated by loops (Figures 12-10b and c). Hydrophobic contacts between α -helices of histone folds are critical for the specific pairing of H2A with H2B and H3 with H4. Histone-fold extensions also make a significant contribution to the specificity of histone pairing. Finally, as the name suggests, flexible tails are located at the ends of histone proteins (Figure 12-10d). The tails are largely unstructured and are involved in interactions with non-histone proteins as well as neighboring nucleosomes.

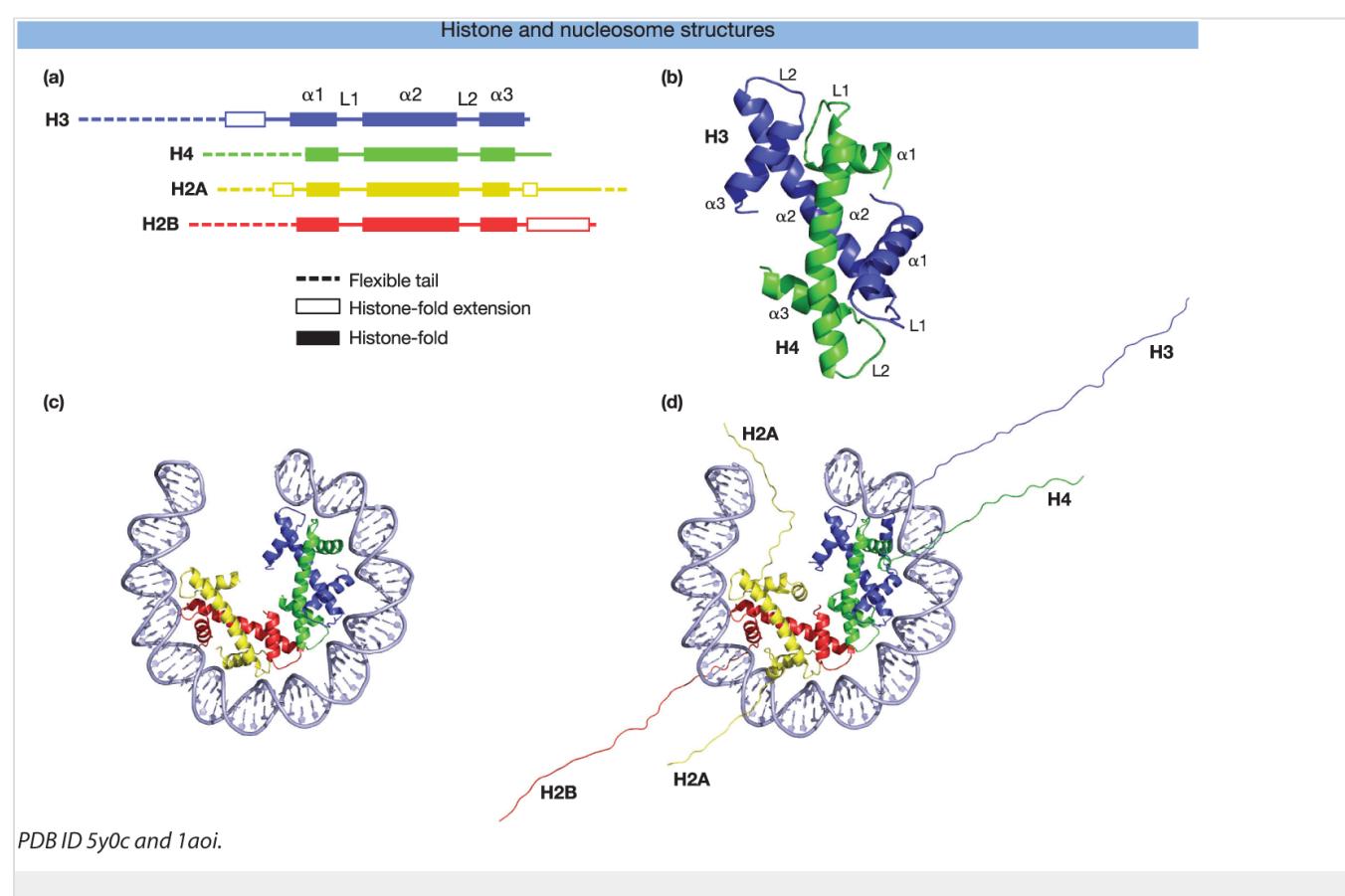


FIGURE 12-10 (a) Primary structures of the four core histones. $\alpha 1$, $\alpha 2$, and $\alpha 3$ are alpha-helices and L1 and L2 are loops. (b) Interactions between the histone-fold structures of histones H3 and H4. (c) The structure of a nucleosome, showing the histone-fold regions of one copy of each of the four core histones wrapped by DNA (gray). (d) The same view of the nucleosome as in (c) with the addition of H2A and H3 histone-fold extensions and flexible tails for all four core histones that extend beyond the wrapped DNA.

The structure of linker histone H1 is substantially different than that of core histones. It is larger (~21 kDa) and it has much greater sequence and structural diversity. For example, in humans, histone H1 has three domains, a central domain of approximately 80 amino acids flanked by unstructured N- and C-terminal tails of approximately 20 and 100 amino acids, respectively; whereas in yeast, histone H1 has only a single, unstructured domain.

Nucleosomes

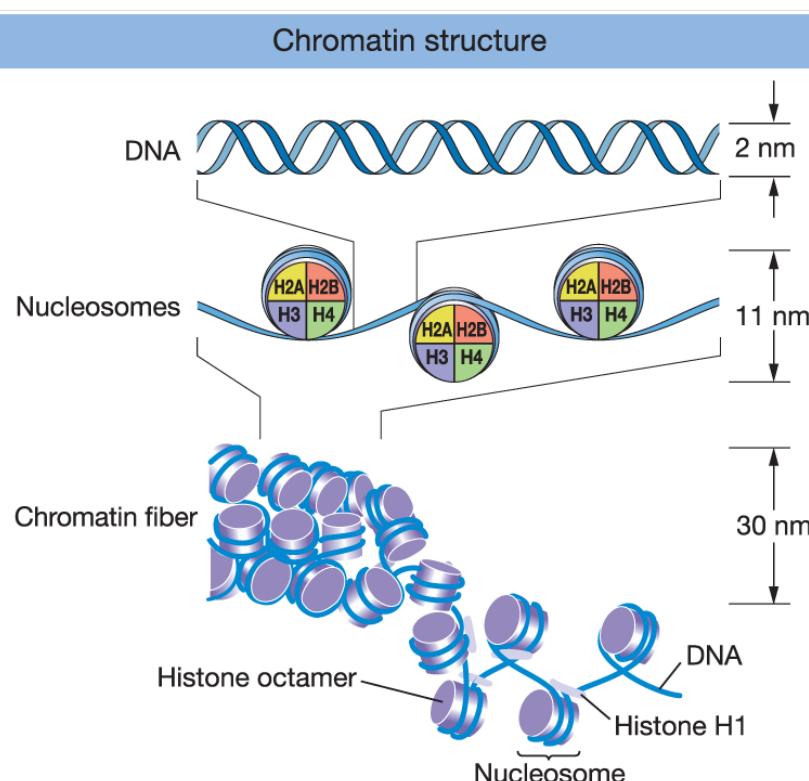
Nucleosomes are the basic structural units of chromatin. They contain 146 base pairs of DNA that wrap about 1.7 times around a **histone octamer** (eight proteins) consisting of two copies of each of the four core histones H2A, H2B, H3, and H4 (only one copy is shown in [Figure 12-10c](#)). The stability of nucleosomes is due to many protein-protein interactions within the histone octamer and electrostatic and hydrogen bonds between histones and DNA. The flexible tails of histones extend away from the nucleosomal DNA and are involved in interactions with adjacent nucleosomes and numerous nuclear factors ([Figure 12-10d](#)).

Molecular machines that assemble and disassemble nucleosomes play important roles in regulating transcription. During DNA replication, nucleosome formation begins with assembly of an H3/H4 tetramer (two H3/H4 dimers joined together) on DNA followed by sequential addition of two H2A/H2B dimers. Binding of histone H1 to nucleosomes organizes an additional 20 base pairs of linker DNA to form a complete nucleosome. Removal of histones from DNA occurs in the reverse order, beginning with sequential removal of the H2A/H2B dimers, followed by removal of the H3/H4 tetramer. Neighboring nucleosomes are separated from one another by ~20–75 base pairs of linker DNA.

KEY CONCEPT In eukaryotes, DNA is packaged with histones in chromatin. Nucleosomes, the units of chromatin, contain two copies of each of the core histones (H2A, H2B, H3, and H4) around which is wrapped 146 base pairs of DNA. Complete nucleosomes also contain histone H1 and linker DNA of variable length.

Chromatin folding

Wrapping of DNA around histone octamers forms a structure of ~11 nanometers (nm) in diameter and compacts DNA ~sixfold ([Figure 12-11](#)). This does not come close to the 10,000-fold compaction that occurs in eukaryotic cells. To achieve higher levels of compaction, nucleosomes fold upon themselves. The next order of chromatin folding produces the 30-nm fiber, a structure ~30 nm in diameter, and there are even more compact structures.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 12-11 Chromatin is made up of 11 nm nucleosomes that fold upon one another into a compact filament of 30 nm.

ANIMATED ART Sapling Plus

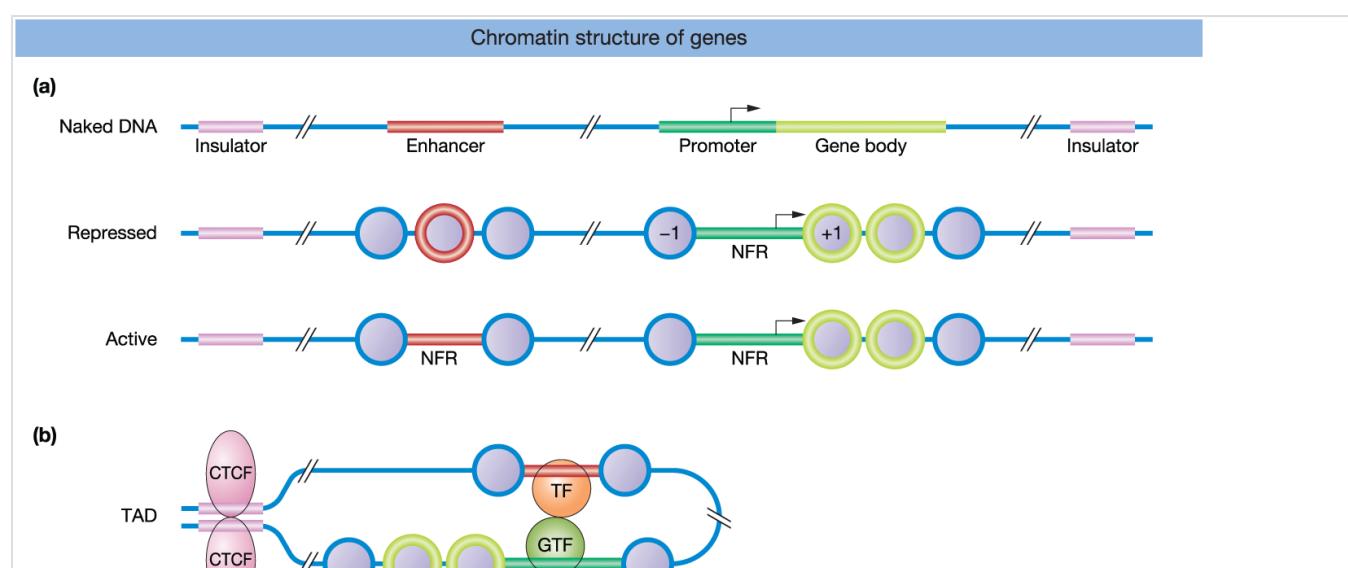
Three-dimensional structure of nuclear chromosomes

During the cell cycle, chromosomes vary in their level of compaction. Nucleosomes in mitosis are much more highly compacted than in interphase (see [Appendix 2-1](#)). Even in interphase, regions of chromosomes vary in their level of compaction. More compacted regions are called **heterochromatin**, and less compacted regions are called **euchromatin**. Heterochromatin constitutes a significant fraction of some eukaryotic genomes—approximately 20 percent for

humans and 30 percent for *Drosophila*—but very little of others—less than 1 percent for the yeast. Chromatin that remains heterochromatic throughout the cell cycle is called **constitutive heterochromatin**, is concentrated at centromeres and telomeres, and is rich in repetitive sequences such as transposons but poor in genes. In contrast, **facultative heterochromatin** is spread along chromosome arms, is gene-rich, and, through mechanisms described later in this chapter, can lose its compact structure and become transcriptionally active euchromatin.

KEY CONCEPT Regions of the genome with few genes, such as centromeres and telomeres, are compacted into heterochromatin throughout the cell cycle, whereas regions that are gene-rich vary in their level of chromatin compaction. Typically, genes are transcriptionally silent when compacted into heterochromatin, and they can be transcriptionally active when less compacted into euchromatin.

Over the past 20 years, new technologies such as chromatin immunoprecipitation (ChIP, [Chapter 14](#)) have made it possible to determine the genome-wide distribution of nucleosomes. As an example, in the yeast *S. cerevisiae*, Frank Pugh and colleagues found that approximately 70,000 nucleosomes occupy 81 percent of the genome and are typically separated by an 18-base-pair linker. Furthermore, nucleosomes are not equally distributed in the genome; they cover 87 percent of transcribed regions, but only 53 percent of intergenic regions. Focusing in on individual genes, transcription start sites are often located within a 150-base-pair **nucleosome free region (NFR)** that contains the promoter and is flanked by positioned nucleosomes, termed the -1 and $+1$ nucleosomes ([Figure 12-12a](#)). The precise positioning of nucleosomes is gradually reduced further upstream and downstream of the promoter. Enhancers are also flanked by a pair of nucleosomes. Enhancers of transcriptionally repressed genes can have nucleosomes positioned at the binding sites for transcription factors, but these nucleosomes are eliminated upon transcription activation.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-12 Genes contain insulator, enhancer, and promoter elements that regulate transcription. (a) Nucleosomes, indicated by purple circles, cover most of the transcribed region of genes (i.e., the gene bodies) but are excluded from enhancers and promoters of active genes to create nucleosome free regions (NFRs). (b) In chromosomes, groups of genes are segregated from one another by topologically associating domains (TADs) that are formed by the interaction of insulator binding proteins such as CTCF. Within TADs, enhancers bound by transcription factors (TFs) are positioned to act on general transcription factors (GTFs) at particular promoters.

The three-dimensional organization of chromatin in the nucleus is not random. Individual chromosomes occupy distinct territories, with gene-dense chromosomes located near the center of the nucleus and gene-poor chromosomes located near the nuclear periphery. Within and between chromosomes, large domains of transcriptionally active chromatin associate with one another. Similarly, inactive chromatin domains associate with one another. Smaller regions of chromatin are organized into **topologically associating domains (TADs)** whose DNA sequences preferentially contact one another. For example, interactions between gene enhancers and promoters are mostly limited to within a TAD ([Figure 12-12b](#)). Anchor points for the looping out of chromatin in TADs are defined by specialized regulatory sequences called **insulators** or **boundaries** that interact with one another or possibly the nuclear envelope through their associated proteins. In mammals, most of the known insulator sequences are bound by a zinc-finger DNA-binding protein called CTCF (CCCTC-binding factor). Therefore, insulators divide chromosomes into precisely defined loops that determine which enhancer-promoter interactions are allowed and which are prevented.

KEY CONCEPT The wrapping of DNA enhancer elements into nucleosomes can prevent binding by transcription factors. Insulators prevent enhancers and their associated transcription factors from activating the transcription of genes outside a TAD.

12.3 CHROMATIN REGULATES TRANSCRIPTION

LO 12.3 Compare and contrast how chromatin modifying and chromatin remodeling mechanisms contribute to gene-specific transcription regulation.

The packaging of eukaryotic DNA into chromatin means that much of DNA is not readily accessible to the transcription machinery. Thus, eukaryotic genes are generally inaccessible and transcriptionally silent unless activated. Two major mechanisms operate in eukaryotic cells to enable dynamic access of the transcription machinery to DNA, resulting in a wide range of transcription states, from silent to highly active.

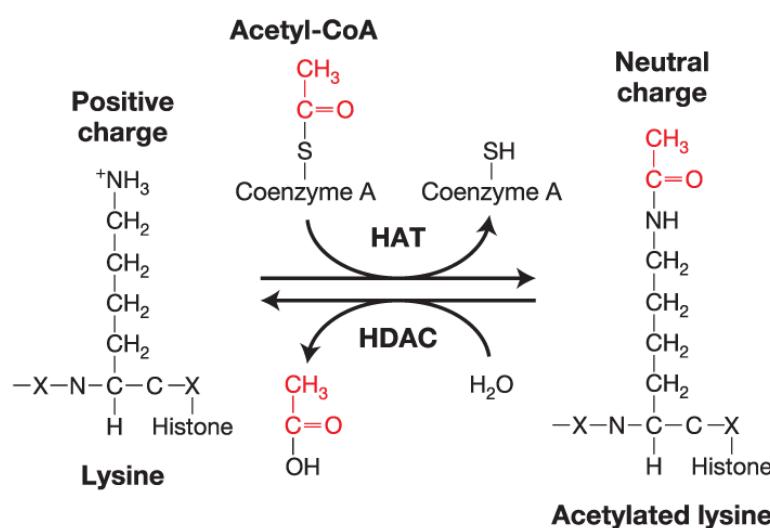
1. In a mechanism called **chromatin modification**, enzymes alter the chemical structure of amino acids in histones or nucleotides in DNA to affect recruitment of transcription factors, coregulators, and general transcription factors to chromatin.
2. In a mechanism called **chromatin remodeling**, the accessibility of DNA to transcription factors, coregulators, and general transcription factors is altered by enzymes that use energy from ATP hydrolysis to remodel nucleosomes; that is, reposition histone octamers along the DNA, remove histone octamers from DNA, or replace canonical histones in octamers with variant histones.

Histone modification: a type of chromatin modification

In 1964, Vincent Allfrey discovered that histones are found in both acetylated and non-acetylated forms. Acetylation is a **post-translational modification** (i.e., it occurs after translation) and consists of addition of an acetyl group to the amino group of a lysine amino acid side chain (**Figure 12-13**). Allfrey hypothesized that histone acetylation affects transcription. His thinking was that acetylation neutralizes the positive charge of lysine and thereby decreases the affinity of lysine for the negatively charged phosphate backbone of DNA. As a consequence, lysine acetylation (i.e., addition of an acetyl group) would reduce chromatin compaction, increase accessibility of the transcription machinery to DNA, and promote transcription activation. Conversely, lysine deacetylation (i.e., removal of the acetyl group) would increase chromatin

compaction, reduce accessibility of the transcription machinery to DNA, and promote transcription repression.

Histone acetylation and deacetylation reactions



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 12-13 The positively charged side chain of lysine residues in histones is neutralized by post-translational addition of an acetyl group. Acetylation is catalyzed by histone acetyltransferases (HATs) that use acetyl-CoA as the donor of the acetyl group, indicated in red. The reverse reaction, deacetylation, is catalyzed by histone deacetylases (HDACs).

Evidence supporting this hypothesis was uncovered in 1996 when David Allis and his colleagues identified the first **histone acetyltransferase (HAT)**, an enzyme that transfers an acetyl group from acetyl-CoA to lysines in histones (Figure 12-13). The HAT called p55 that they identified in Tetrahymena (a ciliated protozoan) turned out to be similar in sequence to a yeast protein called GCN5 that functions as a transcription coactivator. GCN5 promotes transcription but it does not directly bind DNA. Thus, GCN5 provides a mechanistic link between histone acetylation and transcription activation. Subsequently, enzymes called **histone deacetylases (HDACs)** were found to repress transcription by removing acetyl groups from lysines in histones.

Acetylation of lysine residues affects transcription by two mechanisms. First, as hypothesized by Allfrey, acetylation leads to more open chromatin by loosening interactions between histones and DNA as well as interactions between nearby nucleosomes. Second, acetylation creates a binding site for a protein motif called a bromodomain. Several transcription regulatory factors, including the TAF1 subunit of the general transcription factor TFIID, contain bromodomains that increase

the affinity of the factor for particular genes by binding acetylated histones. Using the nomenclature introduced in [Chapters 8](#) and [9](#), HATs are writers, HDACs are erasers, and bromodomains are readers of histone acetylation ([Figure 12-14](#)).

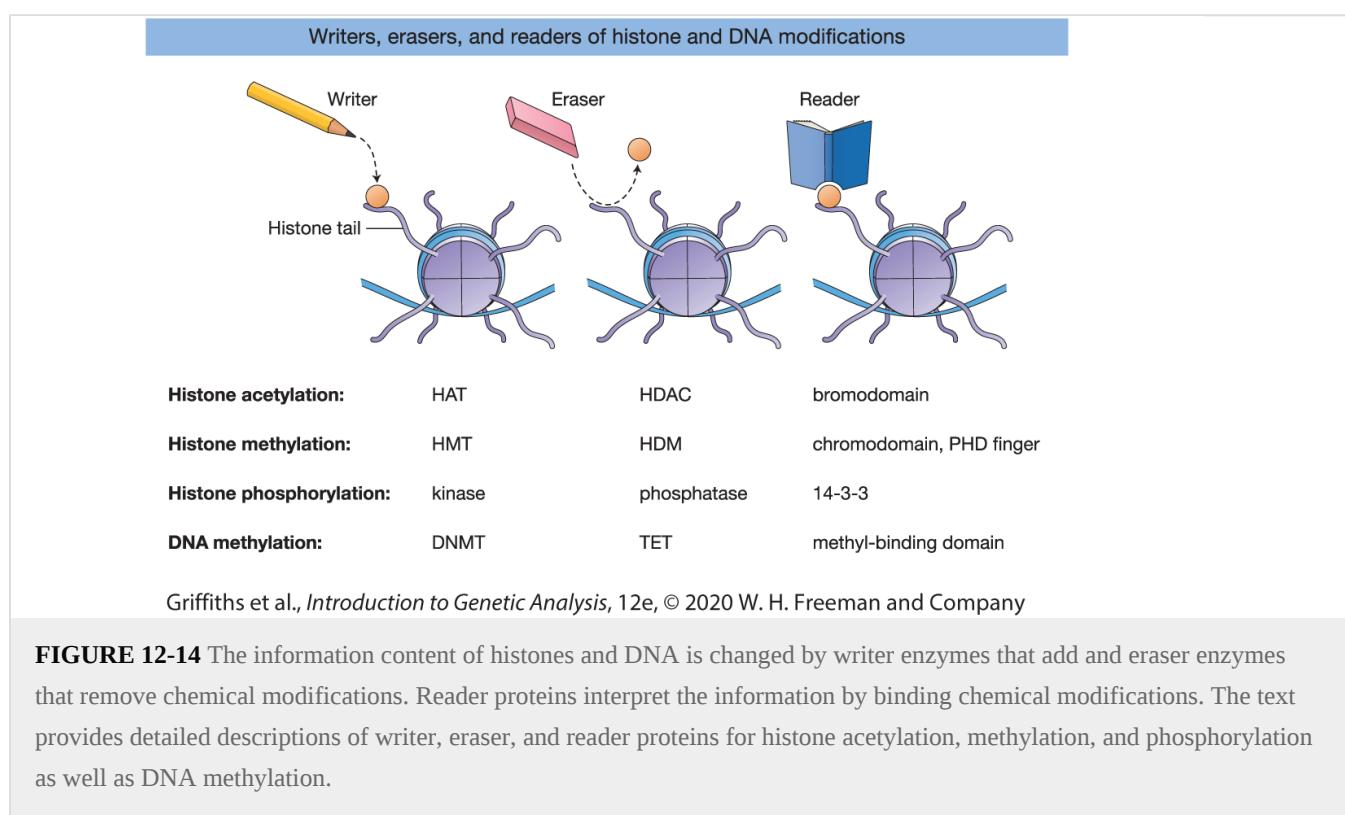


FIGURE 12-14 The information content of histones and DNA is changed by writer enzymes that add and eraser enzymes that remove chemical modifications. Reader proteins interpret the information by binding chemical modifications. The text provides detailed descriptions of writer, eraser, and reader proteins for histone acetylation, methylation, and phosphorylation as well as DNA methylation.

KEY CONCEPT Acetylation of lysines in histones by HATs (1) loosens interactions within and between nucleosomes and (2) creates a binding site for bromodomains, found in some transcription coregulators.

In summary, histone acetylation plays a crucial role in stepwise mechanisms that activate transcription: a transcription factor binds an enhancer, a HAT such as GCN5 binds the transcription factor, the HAT acetylates histones in nucleosomes at the promoter, a bromodomain protein such as TAF1 binds acetylated histones, and RNA polymerase II is recruited either directly or indirectly by the bromodomain protein. Similarly, an activated gene is turned off by transcription factor-mediated binding of an HDAC, which deacetylates histones and blocks recruitment of bromodomain-containing proteins. Acetylation affects transcription initiation, and elongation, by being targeted to nucleosomes positioned in different regions of genes and thereby affecting recruitment of bromodomain-containing initiation and elongation factors ([Chapter 8](#)).

Acetylation is one of many [histone modifications](#) that affect transcription. Other abundant modifications include methylation of lysine and arginine residues; phosphorylation of serine, threonine, and tyrosine residues; and ubiquitination of lysine residues. Furthermore, lysine can be

methylated one, two, or three times (monomethyllysine, dimethyllysine, and trimethyllysine, respectively); while arginine can be methylated one time (monomethylarginine) or two times in symmetric or asymmetric configurations (dimethylarginine) (**Figure 12-15**). Methylation is controlled by writers (histone methyltransferases, HMTs), erasers (histone demethylases, HDMs), and readers (proteins that contain a chromodomain or a plant homeodomain (PHD) finger). There are also writers, erasers, and readers for phosphorylation, ubiquitination, and other modifications, and both histones and DNA can be modified.

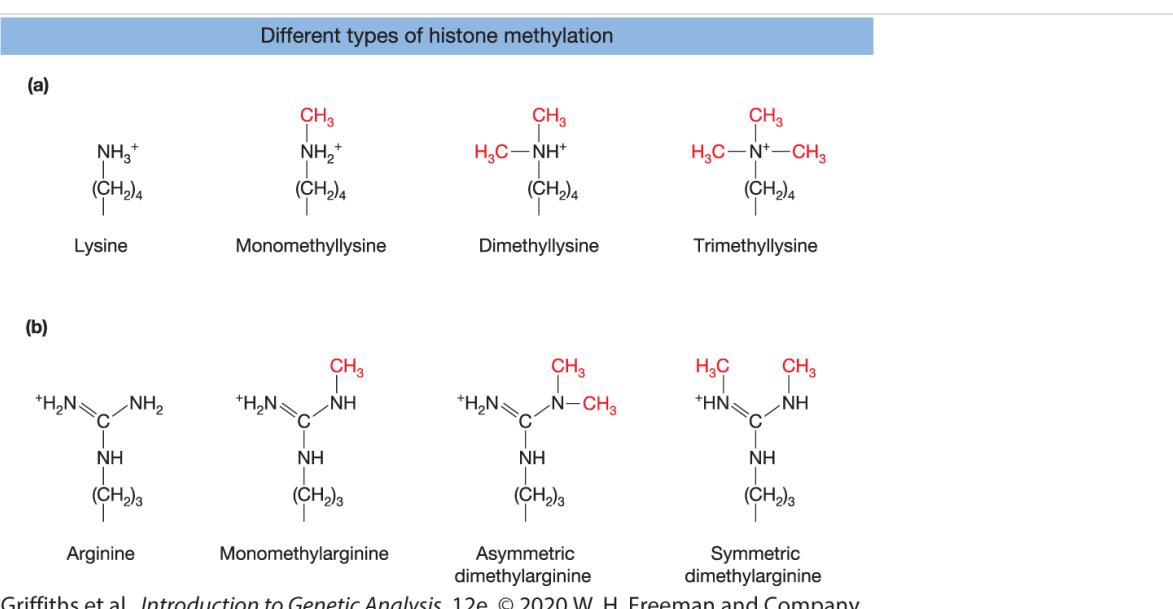


FIGURE 12-15 There are several types of lysine and arginine methylation, each of which conveys different instructions to the transcription regulatory machinery. (a) Different types of lysine methylation. (b) Different types of arginine methylation. Only the amino acid side chain is drawn, with methyl groups shown in red.

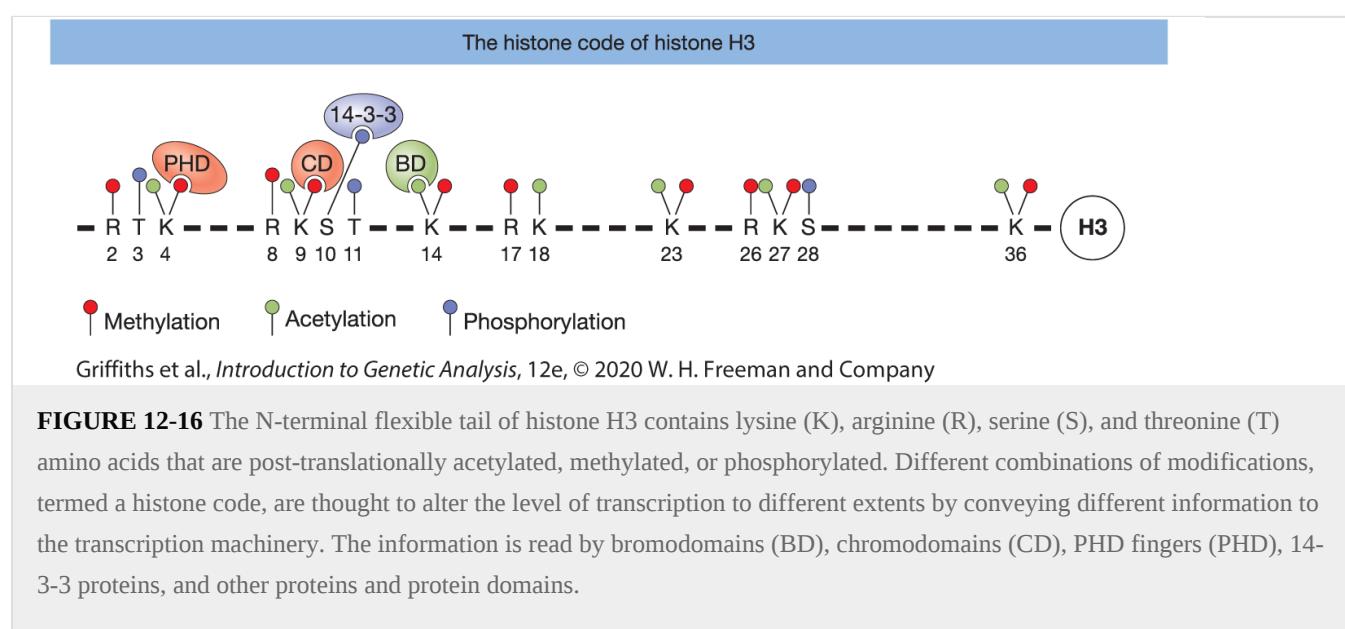
KEY CONCEPT Transcription is regulated by chemical modifications of amino acids in histones and nucleotides in DNA. Modifications are added by writer enzymes, removed by eraser enzymes, and bound by reader proteins.

The histone code hypothesis

Post-translational modifications occur in all parts of histone proteins but are concentrated in the tails. They are experimentally detected *in vivo* and *in vitro* using modification-specific antibodies, and *in vitro* by mass spectrometry of histones purified from cells. Unfortunately, these methods are largely unable to detect the extent to which modifications coexist on an individual histone protein. This information may be very important because different combinations of histone

modifications may convey the information to bring about different transcription outputs. In 2000, this idea was formalized by David Allis and Thomas Jenuwein in the **histone code** hypothesis, which proposes that multiple histone modifications, acting sequentially or in combination on one or several histone tails, specify unique transcription outcomes.

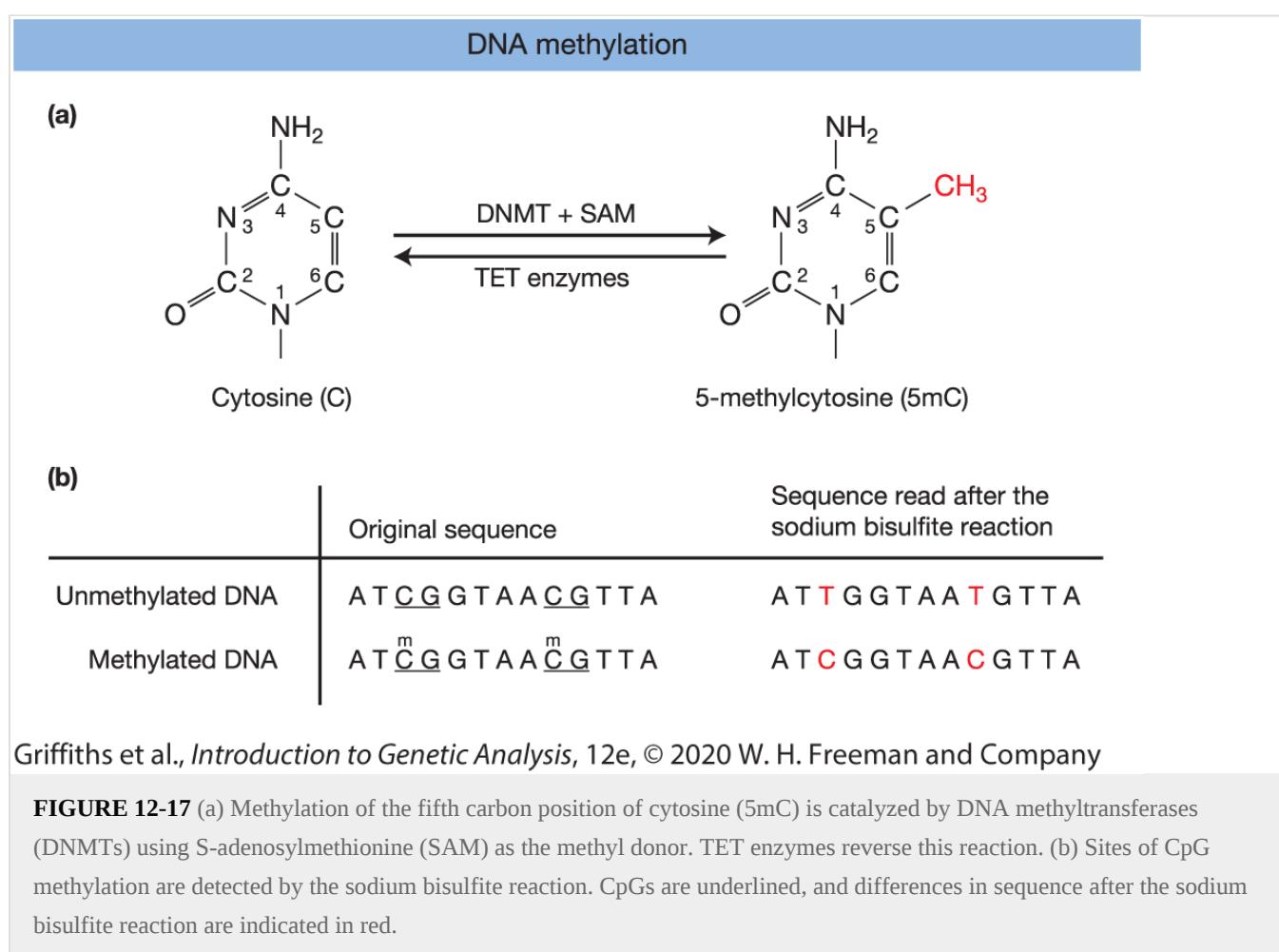
In support of the histone code hypothesis, nucleosomes at promoters of transcriptionally active genes are commonly trimethylated (me3) on lysine (K) 4 of histone H3 (H3K4me3), whereas promoters of transcriptionally repressed genes are trimethylated on H3K9 (H3K9me3) ([Figure 12-16](#)). H3K4me3 activates transcription by serving as a binding site for transcription coactivators such as the PHD finger-containing TAF3 subunit of TFIID. In contrast, H3K9me3 represses transcription by serving as a binding site for transcription corepressors such as the chromodomain-containing protein heterochromatin protein 1 (HP1), which promotes the formation of heterochromatin. Combinations of modifications are also hallmarks of transcription activity—the combination of phosphorylation (P) of serine (S) 10 on histone H3 (H3S10P), which is bound by 14-3-3 proteins, and acetylation of lysine 14 on histone H3 (H3K14ac), which is bound by bromodomain proteins, signals transcription activation. The potential information content of histone modifications is enormous. For example, there are more than two million possible combinations of modifications that can occur on the N-terminal tail of histone H3.



KEY CONCEPT The histone code hypothesis posits that different combinations of histone modifications create unique binding sites that can be read by transcription coregulators, thereby conferring a variety of transcriptional outcomes.

DNA modification: another type of chromatin modification

Like histone modifications, **DNA modifications** affect transcription. In vertebrates, the predominant DNA modification is 5-methylcytosine (5mC), where methylation by a DNA methyltransferase (DNMT) occurs at the fifth carbon in the cytosine ring of the dinucleotide CpG (cytidine-phosphodiester bond-guanosine) (**Figure 12-17a**). DNMTs use S-adenosyl methionine (SAM) as the methyl donor. In contrast to vertebrates, in plants, 5mC occurs in the dinucleotide CpG as well as other nucleotide contexts, and in *Drosophila*, *C. elegans*, and *S. cerevisiae*, little or no DNA methylation has yet been detected.

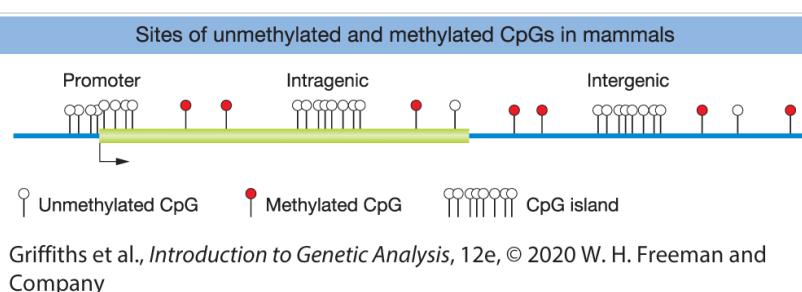


 CpG methylation is reversed by enzymes in the TET (ten-eleven translocation) family (**Figure 12-17a**). Reversal comprises three steps, with each step catalyzed by a TET enzyme; the second and third intermediates are rapidly excised by a mechanism described in **Chapter 15**. In contrast, the

first intermediate is more stable and is particularly abundant in embryonic stem cells and adult neurons where it is bound by reader proteins that regulate genes involved in development and tumorigenesis.

5mC is detected in the lab by the sodium bisulfite reaction. DNA isolated from cells is treated with sodium bisulfite, which in single-stranded DNA efficiently converts cytosine to uracil but inefficiently converts 5mC to thymine. After the sodium bisulfite reaction, the DNA is sequenced, and cytosines that are converted to uracil are read as thymine (T), whereas 5mC that are unchanged are read as cytosine (C) ([Figure 12-17b](#)).

In vertebrates, CpGs occur much less frequently than would be expected based on the C + G content of genomes. This is due to widespread methylation of CpGs and subsequent conversion by deamination over evolutionary time to TpG. Approximately 85 percent of CpGs are methylated and are scattered throughout the genome ([Figure 12-18](#)). The remaining CpGs are unmethylated, and many of these are highly clustered in 200- to 4000-base-pair regions called **CpG islands**. Approximately half of all CpG islands are located in gene promoters, and the remaining half are roughly equally divided between intragenic and intergenic locations. The majority of gene promoters are associated with a CpG island.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-18 Mammalian genomes contain clusters of CpGs called CpG islands as well as dispersed CpGs that can be unmethylated or methylated. A representative gene is indicated by the green bar.

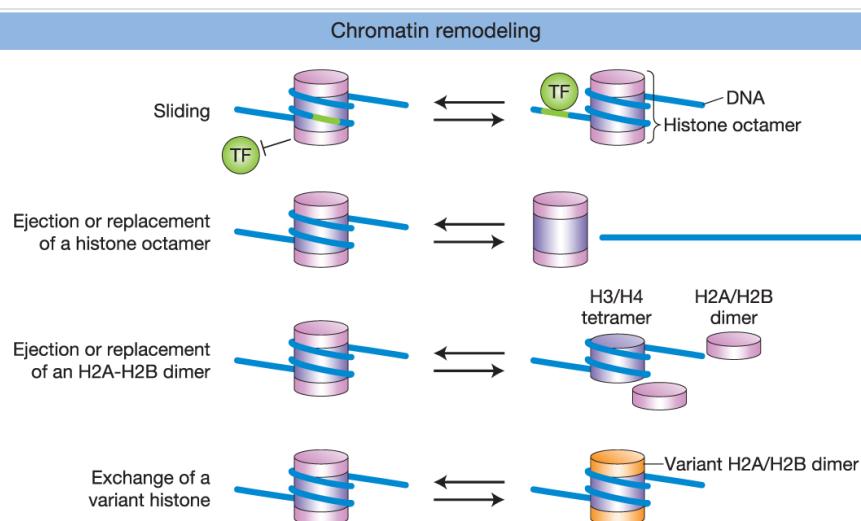
A Unmethylated CpG islands at promoters are generally correlated with open chromatin and active transcription, whereas methylated islands are associated with closed chromatin and repressed transcription. These effects on transcription are mediated by proteins that distinctly bind unmethylated or methylated CpGs. In humans, the protein Cfp1 (CxxC finger protein 1) binds unmethylated CpGs and recruits a histone methyltransferase (HMT) that produces the transcription activating histone modification H3K4me3. In contrast, methylation of CpGs represses transcription by interfering with transcription factor binding to enhancers and by serving as a binding site for methyl binding domain (MBD) proteins that recruit transcription repressors such as HDACs that deacetylate lysine or HMTs that produce H3K9me3. Thus, DNMTs are

writers, TETs are erasers, and a variety of proteins are readers of unmethylated and methylated CpGs.

KEY CONCEPT Methylation of cytosine in CpG islands at gene promoters is correlated with the repression of transcription. Like modifications of histone proteins, CpG methylation of DNA represses transcription by altering the affinity of transcription factors, coregulators, and general transcription factors for chromatin.

Chromatin remodeling

Chromatin remodeling is the workhorse in the process of altering chromatin structure to regulate transcription. Remodeling means changing histone-DNA interactions in nucleosomes to render DNA either more or less accessible to transcription regulators ([Figure 12-19](#)). To permit factors to bind enhancer and promoter elements, remodeling moves nucleosomes to new locations. On moderately active genes, remodeling displaces H2A/H2B dimers in front of RNA polymerase II and replaces them behind the polymerase during every round of transcription. H3/H4 tetramers may only be displaced and replaced on highly active genes. In contrast, eviction of histones from DNA may not be necessary to permit RNA polymerase II passage through genes being transcribed at a low level.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-19 Chromatin remodeling complexes use energy produced by ATP hydrolysis to slide, eject, or replace histone octamers on DNA or exchange variant histones for canonical histones in octamers. TF indicates a transcription factor that binds the green enhancer element.

ANIMATED ART  Sapling Plus

Chromatin remodeling

Chromatin remodeling complexes use energy from ATP hydrolysis to disrupt non-covalent histone-DNA interactions. Eukaryotic organisms contain four families of ATP-dependent chromatin remodeling complexes, which can be characterized as being associated with transcription activation or repression.

Two genetic screens in yeast for mutants in seemingly unrelated processes led to the discovery of a chromatin remodeling complex. In one study, mutagenized yeast cells were screened for the lack of growth on sucrose (sugar *nonfermenting* mutants, *snf*, pronounced “sniff”). In the other study, mutagenized yeast cells were screened for defective switching of mating type (*switch* mutants, *swi*, pronounced “switch”). Many mutants for different loci were recovered in each screen, but one mutant gene was found to cause both phenotypes. Mutants at the *swi2/snf2* locus could neither use sucrose effectively nor switch mating type because the transcription of specific genes was blocked. The protein encoded in the *swi2/snf2* locus was found to be the ATPase subunit of the multisubunit SWI/SNF (“switch-sniff”) chromatin remodeling complex.

SWI/SNF affects transcription by remodeling nucleosomes in two steps, initially removing an H2A/H2B dimer from DNA, followed by removal of the rest of the histone octamer. The gene specificity of the SWI/SNF complex is provided by binding to transcription factors such as Gal4 and through binding of a bromodomain-containing subunit of the complex to acetylated lysine. Thus, transcription factors, histone modifying enzymes, and chromatin remodeling factors function in concert to regulate transcription.

In contrast to the other chromatin remodeling complexes that slide, eject, or replace histone octamers on DNA, the SWR1 (pronounced “swur one”) complex remodels chromatin by assembling the variant histone H2A-Z into chromatin. SWR1 does this by exchanging H2A-Z/H2B dimers for H2A/H2B dimers in histone octamers. This activity is targeted to enhancers and promoters of specific genes by the bromodomain-containing subunit of SWR1, which binds specific acetylated lysines on histone tails. Nucleosomes that contain H2A-Z are particularly prone to disassembly by other chromatin remodeling complexes, leading to increased access of transcription regulators to DNA.

KEY CONCEPT Chromatin is dynamic; nucleosomes are not necessarily in fixed positions on the chromosome. Chromatin remodeling complexes change nucleosome density, position, and subunit composition to control access of the transcription machinery to DNA.

Connecting chromatin structure to transcription: lessons from the *interferon- β* gene

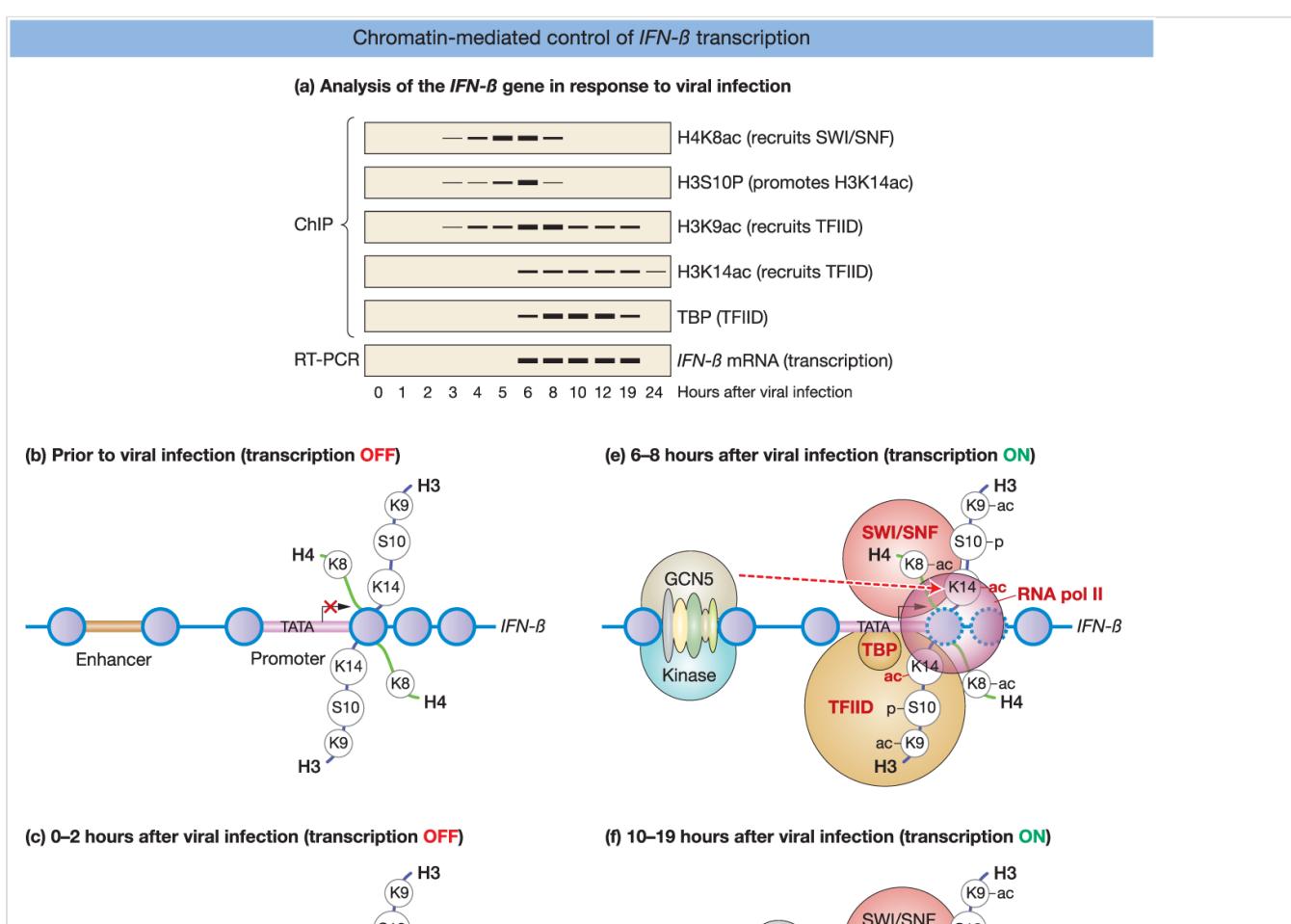
The human *interferon-beta* (*IFN- β*) gene, which encodes an antiviral protein, is one of the best-characterized genes in eukaryotes. Its transcription is normally switched off but, upon viral infection, is activated to very high levels. A central feature of activation of this gene is assembly of multiple, different transcription factors into an **enhanceosome** about 100 base pairs upstream of the TATA box promoter element and transcription start site.

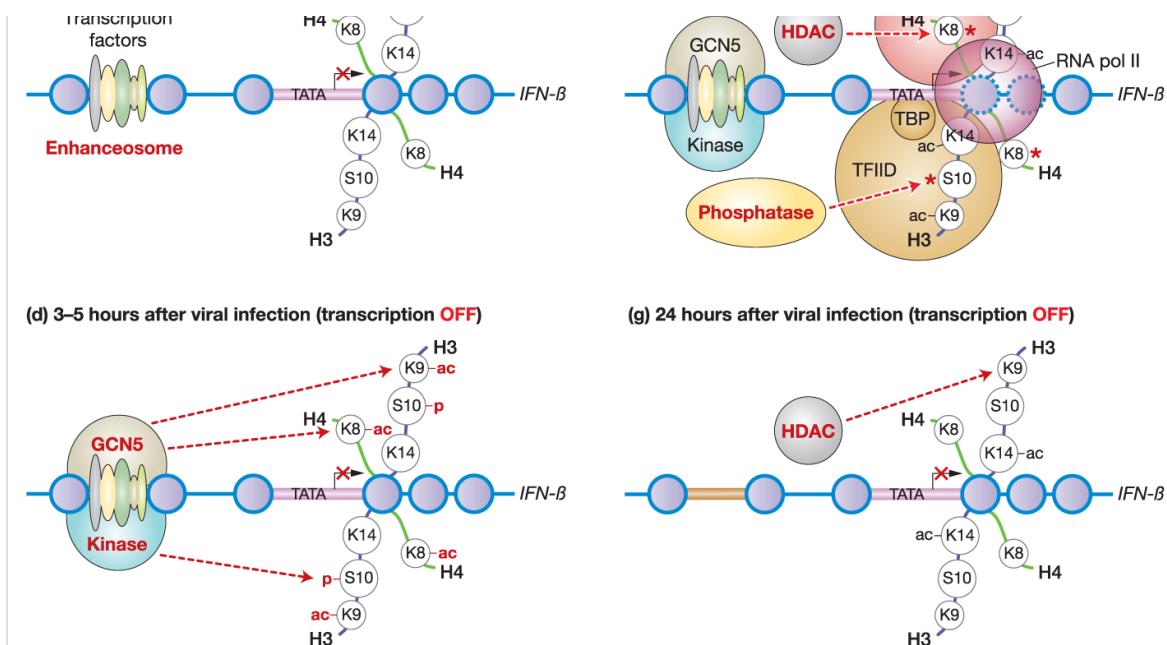
A study of the *IFN- β* gene illustrates how regulated changes in chromatin structure affect transcription. Dimitris Thanos and colleagues used the chromatin immunoprecipitation (ChIP) technique ([Chapter 14](#)) to identify histone modifications and other events that occur at the *IFN- β* promoter as the gene shifts from transcriptionally inactive to active and back to inactive over a 24-hour period in response to virus infection of human cells ([Figure 12-20a](#)). They also used reverse transcription-PCR (RT-PCR) to quantify *IFN- β* mRNA levels. The following bullet points walk through the temporal pathway of molecular events uncovered in the study and they highlight the general mechanisms by which chromatin structure affects transcription:

- *IFN- β* mRNA is first detected 6 hours after viral infection, but histone modifications are detected as early as 3 hours after viral infection ([Figure 12-20a](#)). This illustrates that histone modifications occur prior to the first transcription initiation event to generate a chromatin structure that is conducive to transcription initiation and elongation.
- The earliest event is cooperative binding of a suite of transcription factors to the *IFN- β* proximal enhancer ([Figure 12-20b](#)) to form an enhanceosome ([Figure 12-20c](#)). The histone acetyltransferase GCN5 then binds the assembled enhanceosome and acetylates H4K8 and H3K9 and a kinase binds and phosphorylates H3S10 in nucleosomes near the *IFN- β* promoter ([Figure 12-20d](#)). This illustrates that transcription factors recruit histone modifying enzymes to generate specific histone modifications in particular nucleosomes.
- H4K8ac and H3S10P are last detected at 8 hours (see [Figure 12-20a](#)), even though transcription occurs up to 15 hours. In contrast, H3K9ac is last detected at 19 hours. Unknown HDACs and phosphatases are involved in removing these histone marks. This

illustrates that both writing and erasing of histone modifications as well as the relative timing of histone modifications are important for transcription regulation.

- H3S10P peaks at 6 hours, the time at which H3K14ac is first detected (see [Figure 12-20a](#)), suggesting that H3S10P is required for GCN5 to acetylate H3K14 ([Figure 12-20e](#)). This illustrates that modification of one amino acid can promote or inhibit modification of other amino acids, a process called crosstalk.
- TBP binding is first detected at 6 hours, which is the same time that *IFN-β* transcription starts (see [Figure 12-20a](#)). The TFIID complex is recruited through direct binding of TBP to the TATA promoter element as well as binding of TAF1 bromodomains to H3K9ac and H3K14ac ([Figure 12-20e](#)). In addition, the SWI/SNF chromatin remodeling complex is recruited by binding to H4K8ac as well as interactions with another HAT called CBP (CREB-binding proteins), which replaces GCN5 at the enhanceosome. To initiate *IFN-β* transcription, TFIID recruits RNA polymerase II to the promoter and SWI/SNF remodels nucleosomes to allow the polymerase to initiate transcription. This illustrates that histone modifications aid in assembly of the transcription preinitiation complex as well as factors that remodel chromatin structure.
- Last, transcription is turned off at 24 hours, at which time activating histone modifications have mostly been removed by eraser enzymes ([Figure 12-20g](#)). This illustrates the rapid reversibility of chromatin-mediated control of transcription.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-20 (a) ChIP and RT-PCR analysis of the *IFN-β* gene. For ChIP, the intensity of the black bands is proportional to the amount of modified histone or TBP associated with the *IFN-β* promoter. For RT-PCR, the intensity of the black bands indicates the amount of *IFN-β* mRNA. Descriptions in parentheses indicate the functional consequence of the experimental outcome. (b–g) Models that illustrate the sequential molecular interactions and enzymatic events that occur over time after viral infection. Purple circles indicate nucleosomes. Dotted purple circles indicate nucleosomes altered by chromatin remodeling. Note that transcription is on only 6 to 19 hours after viral infection. [Data from Agalioti et al., “Deciphering the Transcriptional Histone Acetylation Code for a Human Gene,” *Cell* 111, 2002, 381–392.]

Lessons learned from the *IFN-β* gene are generalizable. The molecular mechanisms that alter chromatin structure turn transcription of particular genes on or off in response to developmental and environmental signals by permitting or preventing transcription regulatory proteins access to DNA. The information that controls transcription consists of DNA regulatory sequences (e.g., enhancers, promoters, and insulators), the histone code (e.g., various chemical modifications of histone amino acids), and DNA modifications (e.g., unmodified CpG and 5mC) ([Figure 12-21a](#)). Proteins read the information through physical interactions with the information-containing elements ([Figure 12-21b](#)). DNA sequences are bound by transcription factors via their DNA-binding reader domain. Modified histones and modified DNA are bound by other coregulators that contain a variety of reader domains (e.g., bromodomain, chromodomain, and methyl-binding domain). Once bound to chromatin, reader proteins serve as scaffolds for assembly of enzymes that change chromatin structure ([Figure 12-21c](#)). Reader proteins are either part of a stable complex that contains enzymes (e.g., SWI/SNF) or they recruit enzymes through protein-protein interactions. The enzymes then either edit the information at the gene by modifying histones or DNA, or they change the chromatin structure by sliding nucleosomes, ejecting or replacing histone octamers or parts of octamers, or exchanging variant histones for canonical histones (

Figure 12-21d). Collectively, these events alter the access to DNA by RNA polymerase II, transcription initiation factors (e.g., TFIID), and transcription elongation factors (e.g., P-TEFb) (Figure 12-21e). Lastly, since these mechanisms are reversible, transcription can be rapidly turned on or off in response to developmental and environmental signals.

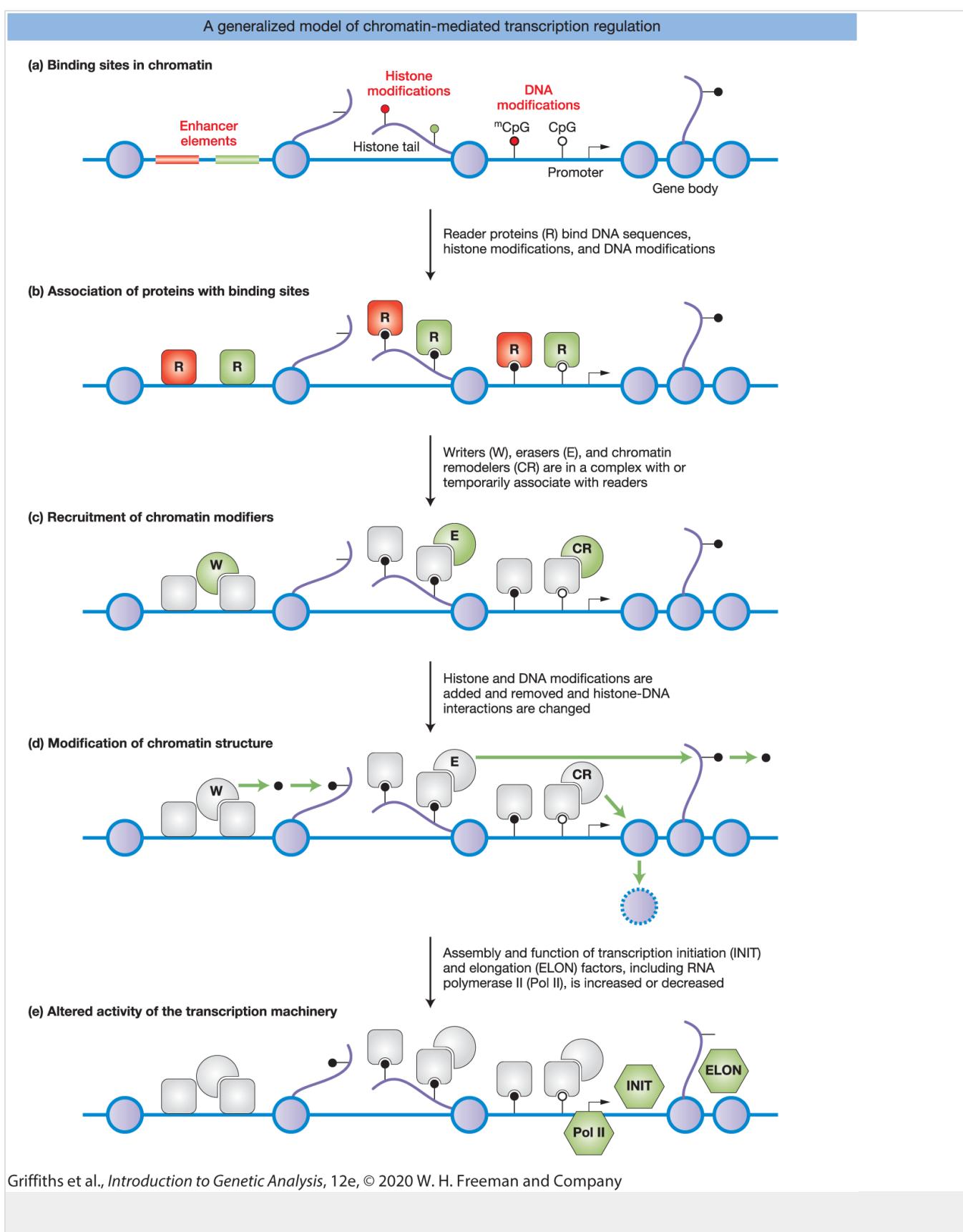


FIGURE 12-21 A model describing how the structural features of chromatin regulate gene-specific transcription in response to signals. Red and green indicate factors that, respectively, activate or repress transcription, and purple circles indicate nucleosomes.

P1 **KEY CONCEPT** *IFN- β* transcription exemplifies how chromatin regulatory strategies are used by cells to alter the transcription of specific genes in response to signals.

12.4 CHROMATIN IN EPIGENETIC REGULATION

LO 12.4 Provide examples of the chromatin-based mechanisms that maintain gene expression over cellular or organismal generations.

When cells divide, information stored in the sequence of DNA is faithfully replicated and transferred to daughter cells. Similarly, information stored in the structure of chromatin is inherited through cell divisions. This form of inheritance is given a special name—**epigenetic inheritance**—because it affects the traits of daughter cells without altering DNA sequence. In this section, we describe four examples of the epigenetic control of transcription: cellular memory, position-effect variegation, genomic imprinting, and X-chromosome inactivation. In each case, the collection of genes that are transcribed in a parent cell is reproduced in daughter cells through the maintenance of chromatin structure by histone and DNA modifying and chromatin remodeling mechanisms.

Cellular memory

Unlike DNA sequence, chromatin structure can change during the life of a cell, and the changes can be inherited in successive generations of cell division. Changes in cell fate are based on short-lived signals that affect the transcription of specific genes. Even after the signal goes away, the cell fate does not change because the effect on transcription stays. For instance, once an embryonic cell differentiates into an intestinal cell, with its intestinal cell-specific spectrum of transcriptionally active and inactive genes, it usually remains an intestinal cell as long as it lives.

Studies that were initially performed in *Drosophila* identified two groups of proteins that function to maintain the cellular memory of transcription, *Polycomb* group proteins and *Trithorax* group proteins. *Polycomb* and *Trithorax* proteins often function in opposition to one another, with *Polycomb* proteins maintaining genes in a transcriptionally *repressed* state and *Trithorax* proteins maintaining genes in a transcriptionally *active* state. Members of the *Polycomb* and *Trithorax* groups are components of multiprotein complexes that post-translationally modify histones and remodel chromatin. For example, a *Polycomb* complex *trimethylates* H3K27 (a histone

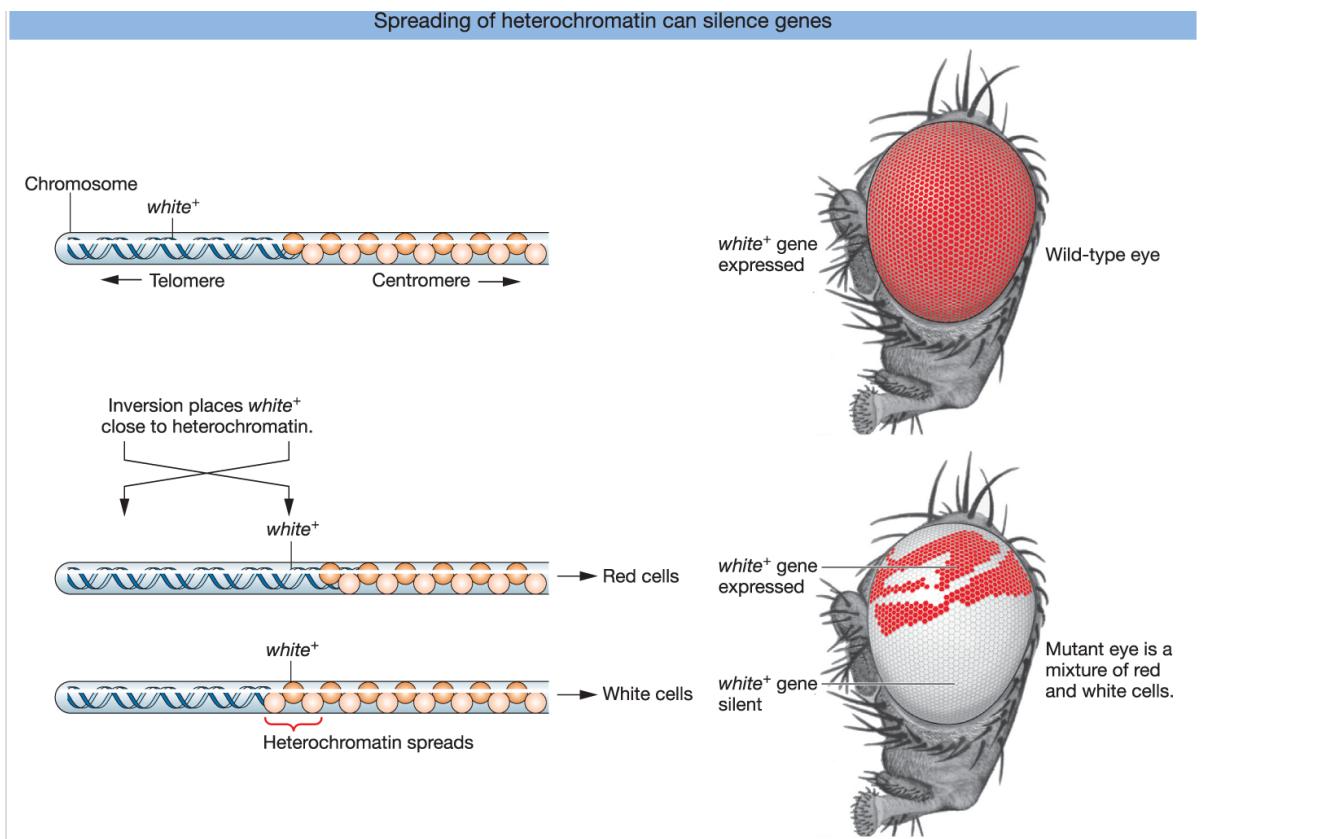
modification commonly associated with transcription silencing), while a Trithorax complex *acetylates* H3K27 (a histone modification commonly associated with transcription activation). Note that since trimethyl-lysine and acetyl-lysine cannot occur at the same time on H3K27, trimethylation by the Polycomb complex blocks the activating acetylation by the Trithorax complex. As with other transcription coregulators, targeting of Polycomb and Trithorax complexes to chromatin is influenced by transcription factors, histone modifications, DNA methylation, and long noncoding RNAs (for example, see X-chromosome inactivation later in this section).

KEY CONCEPT Polycomb and Trithorax group proteins work in opposition to maintain the repressed and active transcription states of parent cells in daughter cells.

Position-effect variegation

In 1930, Hermann Muller discovered an interesting genetic phenomenon while studying *Drosophila*. He found that the expression of genes can be silenced when they are experimentally “relocated” to another region of a chromosome. In these experiments, flies were irradiated with X rays to induce mutations in their germ cells, and the progeny of the irradiated flies were screened for unusual phenotypes. Among the collection of mutants, Muller found flies with eyes that had patches of red and white color. This is unusual because wild-type flies have uniform red eyes, and flies that are mutant for the *white* gene, which is required for the production of red pigment, have uniform white eyes.

Cytological examination revealed a chromosomal rearrangement in the mutant flies: a region of the X-chromosome containing the *white* gene was inverted ([Figure 12-22](#)). Inversions and other chromosomal rearrangements will be discussed in [Chapter 17](#). In this rearrangement, the *white* gene, which is normally located in a euchromatic region of the X-chromosome, now is near the heterochromatic centromere. The patchy eye phenotype of Muller’s flies is due to spreading of heterochromatin into the wild-type *white* gene and silencing of *white* transcription in some cells but not others. Patches of white tissue in the eye are derived from descendants of a single cell in which the *white* gene is silenced and remains silenced through future cell divisions. In contrast, red patches arise from cells in which heterochromatin has not spread into the *white* gene, and so the *white* gene remains active in all its descendants.



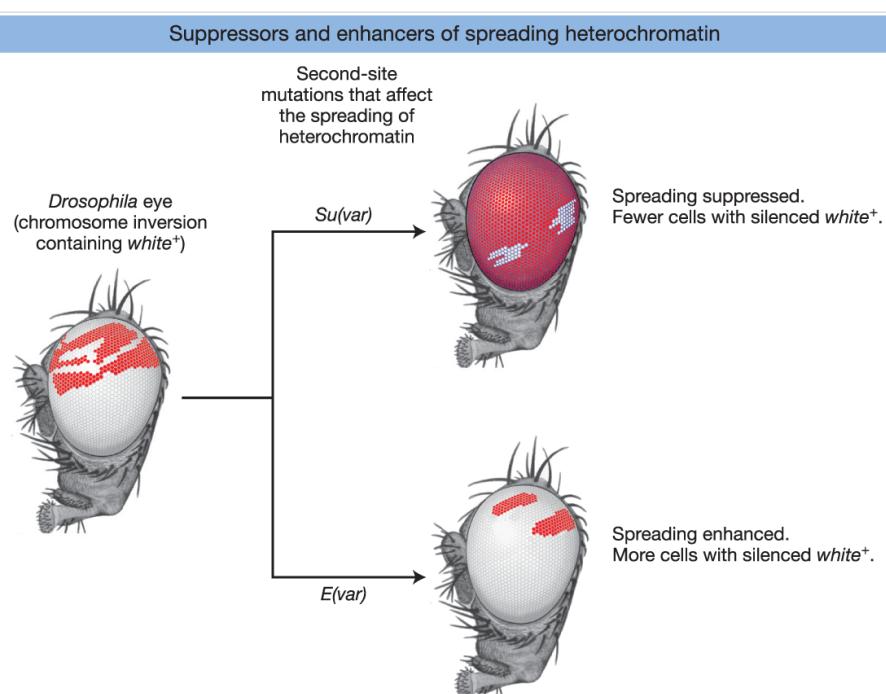
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-22 Chromosomal rearrangement produces position-effect variegation (PEV). A chromosomal inversion places the wild-type *white* gene close to heterochromatin, and spreading of heterochromatin into *white* silences transcription. Wherever *white* is silenced, cells are white because they do not make red pigment.

The existence of red and white patches of cells in the eye of a single organism dramatically illustrates two features of epigenetic transcription regulation. First, as described earlier, differences in chromatin structure across chromosomes can be inherited from one cell generation to the next. Second, differences in chromatin structure across chromosomes affect the expression of resident genes.

Findings from subsequent studies in *Drosophila* and yeast demonstrated that many active genes are silenced in this mosaic fashion when they are relocated to neighborhoods near centromeres or telomeres that are heterochromatic. Furthermore, the effect of local chromatin structure on transcription is not limited to centromeres and telomeres. In mouse cells the degree of chromatin compaction at the site of integration of a transgene correlates with the level of transcription of the transgene and accounts for about a 1000-fold variation in transcription level across the genome. This phenomenon has been called **position-effect variegation (PEV)**. It provides powerful evidence that chromatin structure is able to regulate the expression of genes—in this case, determining whether genes with identical DNA sequence will be active or silent.

Geneticists reasoned that PEV could be exploited to identify the proteins necessary for forming heterochromatin. To this end, they isolated mutations that either suppressed or enhanced the variegated pattern (**Figure 12-23**). A Suppressor of variegation (*Su(var)*) is a gene that when mutated reduces the spread of heterochromatin, meaning that the wild-type product of this gene is required for spreading. In contrast, an Enhancer of variegation (*E(var)*) is a gene that when mutated increases the spread of heterochromatin and normally functions to block spreading. *Su(var)* and *E(var)* genes have proved to be a treasure trove for scientists interested in the proteins that are required to establish and maintain the heterochromatic state.

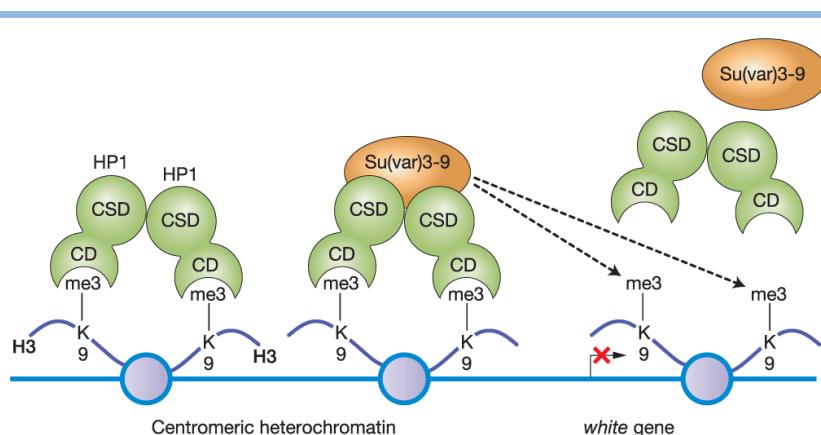


Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-23 Forward genetic screens were used to identify genes that suppress, *Su(var)*, or enhance, *E(var)*, position-effect variegation.

Among more than 300 *Drosophila* mutants identified by these screens was *Su(var)2-5*, which encodes a histone reader protein heterochromatin protein 1 (HP1), and *Su(var)3-9*, which encodes a histone methyltransferase (**Figure 12-24**). HP1 contains a chromodomain that binds H3K9me3 and a chromoshadow domain involved in dimerization of HP1 proteins and recruitment of a variety of chromatin-modifying factors. On the other hand, *Su(var)3-9* trimethylates H3K9. HP1 and *Su(var)3-9* interact with one another to create a feed-forward loop that spreads heterochromatin. HP1 binds H3K9me3 and dimerizes with another HP1 molecule; the dimer recruits *Su(var)3-9*, which generates H3K9me3; and HP1 binds H3K9me3 to continue the process.

A model of spreading heterochromatin



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-24 The coordinated activities of HP1 and *Su(var)3-9* spread heterochromatin from the centromere into the repositioned *white* gene. The dotted arrows indicate that *Su(var)3-9* trimethylates H3K9 on adjacent nucleosomes, recruiting HP1 and silencing transcription of the *white* gene. CD and CSD indicate the HP1 chromodomain and chromoshadow domain, respectively, and purple circles indicate nucleosomes.

In the absence of any barriers, heterochromatin might spread into adjoining regions and inactivate genes in some cells but not in others. One can imagine that the spreading of heterochromatin into active gene regions could be disastrous for an organism because active genes would be silenced as they are converted into heterochromatin. To avert this potential disaster, boundary/insulator elements, which were discussed earlier in the context of topologically associating domains (TADs, Figure 12-12) prevent the spreading of heterochromatin by creating a local environment that is not favorable to heterochromatin formation. Insulator-binding proteins may block the spread of heterochromatin by recruiting activating enzymes such as histone acetyltransferases, H3K4 methyltransferases, and SWI/SNF chromatin remodelers, or they may block access to histones by directly binding them.

KEY CONCEPT Proteins involved in the spread of heterochromatin include writers, readers, and erasers of histone modifications.

Genomic imprinting

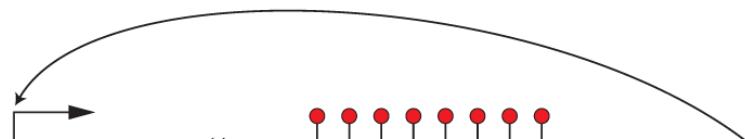
The phenomenon of **genomic imprinting** was discovered about 35 years ago in mammals. In genomic imprinting, certain autosomal genes are expressed in a parent-of-origin-specific manner. For example, transcripts from the *Igf2* (*insulin-like growth factor 2*) gene in mammals come exclusively from the father's (i.e., paternal) allele because the mother's (i.e., maternal) allele is

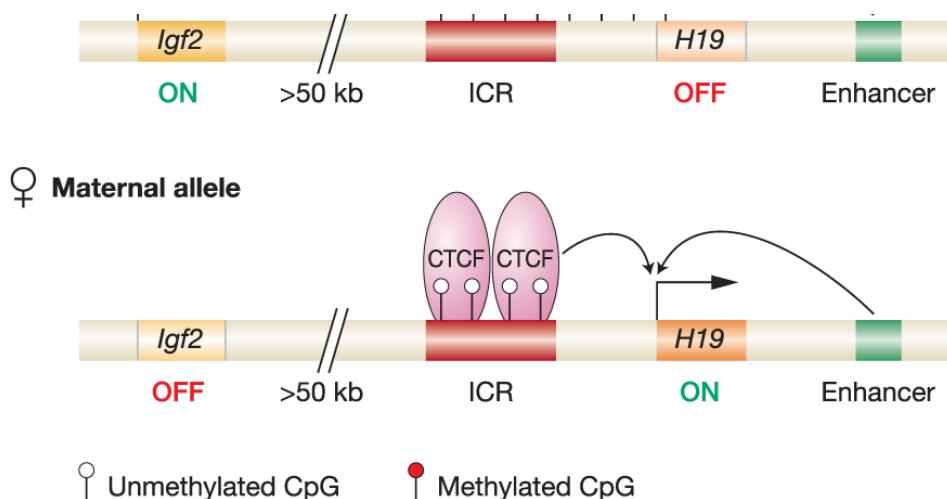
silenced—an example of **maternal imprinting** because the copy of the gene derived from the mother is transcriptionally inactive. Conversely, *H19* transcripts come exclusively from the mother's allele; *H19* is an example of **paternal imprinting** because the paternal copy is transcriptionally inactive. The consequence of parental imprinting is that imprinted genes are expressed as if only one copy of the gene is present in the cell even though there are two. Importantly, no changes are observed in the DNA sequences of imprinted genes; that is, the identical gene can be active or inactive in the progeny, depending on whether it was inherited from the mother or father. Imprinted genes are controlled by DNA regulatory elements called imprinting control regions (ICRs) that have parent-specific chromatin modifications. This then represents an epigenetic phenomenon.

Let's turn again to the mouse *Igf2* and *H19* genes to see how imprinting works at the molecular level. These two genes are located in a cluster of imprinted genes on mouse chromosome 7. There are an estimated 100 imprinted genes in the mouse, and most are found in clusters containing 3–12 imprinted genes that are spread out over 20 kilobases to 3.7 megabases of DNA. Humans have most of the same clustered imprinted genes as mice. In all cases examined, there is a specific pattern of DNA methylation and histone modification at the ICR for each parental copy of an imprinted gene. For the *Igf2*–*H19* cluster, the ICR DNA that lies between the two genes is methylated in male germ cells and unmethylated in female germ cells ([Figure 12-25](#)). Thus, methylation of the ICR leads to *Igf2* being transcriptionally active and *H19* being inactive, whereas the lack of methylation leads to the reverse. This difference is due to the fact that only the unmethylated (female) ICR can be bound by CTCF, the same protein that binds insulator elements in TADs ([Figure 12-12](#)). When bound, CTCF acts as an enhancer-blocking insulator that prevents enhancer activation of *Igf2* transcription. However, the enhancer in females can still activate *H19* transcription. In males, CTCF cannot bind to the ICR, and the enhancer can activate *Igf2* transcription (recall that enhancers can act at great distances). However, the enhancer cannot activate *H19* because the methylated region extends into the *H19* promoter. Epigenetic marks such as DNA methylation that cause genes to be expressed in a parent-of-origin manner are established in germ cells (sperm and eggs) and, as organisms develop, are maintained through mitotic cell division of somatic cells.

A model of genomic imprinting

♂ **Paternal allele**





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

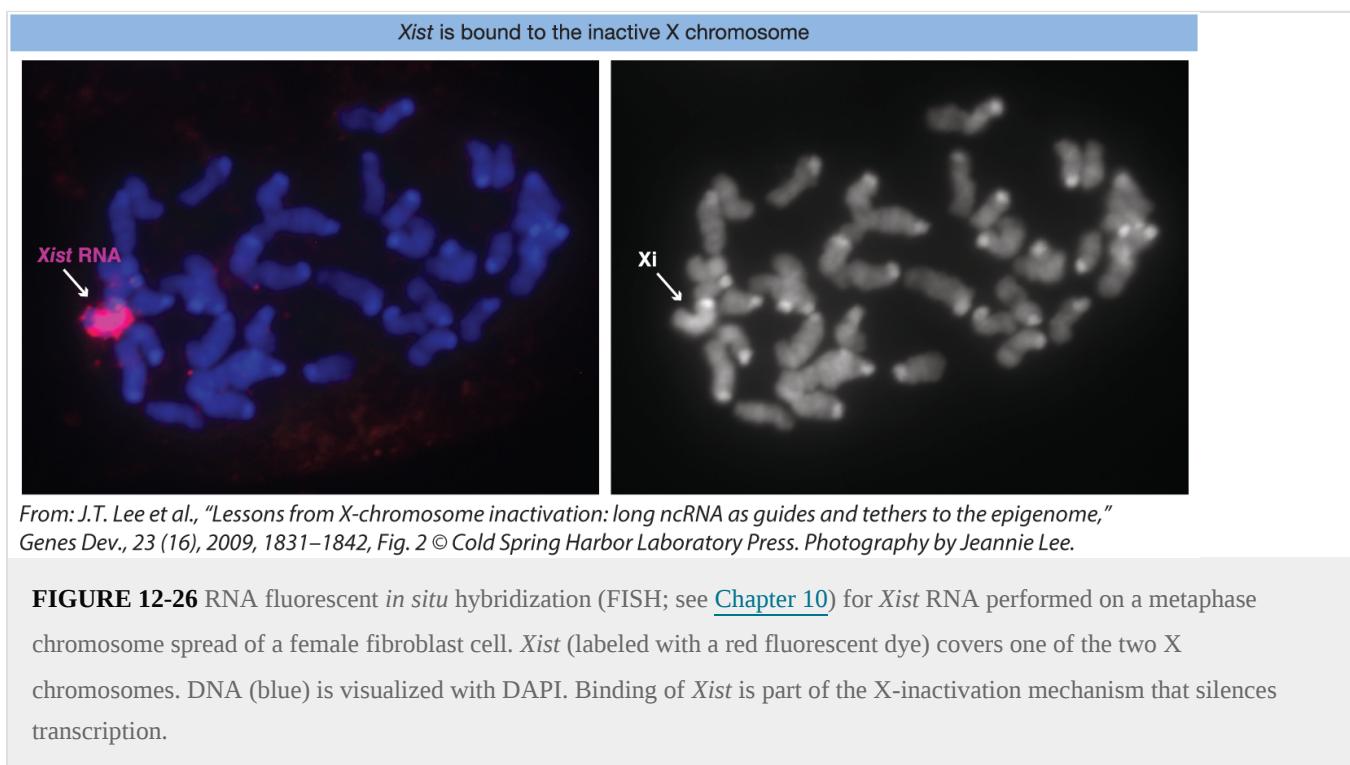
FIGURE 12-25 Genomic imprinting in the mouse. The imprinting control region (ICR) is methylated in male gametes, blocking CTCF binding and directing the enhancer to activate transcription of *Igf2*. The unmethylated ICR in female gametes binds CTCF, forming an insulator that blocks enhancer activation of *Igf2* and directs activation of *H19*.

KEY CONCEPT For most diploid organisms, both alleles of a gene are expressed independently; however, a few genes in mammals undergo genomic imprinting. Through this mechanism, epigenetic marks made in germline cells are retained throughout development of offspring, silencing one allele and allowing expression of the other.

Note that parental imprinting can greatly affect disease inheritance. For most diploid genes, mutation of the copy inherited from one parent does not produce a disease phenotype because there is an additional copy from the other parent. However, imprinted genes are essentially haploid because only one of the two copies is expressed. Thus, as you might expect, diseases occur due to mutations in the non-imprinted, transcriptionally active, copy of imprinted genes. Prader–Willi syndrome and Angelman syndrome are examples of imprinting diseases derived from loss of non-imprinted paternal and maternal genes, respectively. These diseases occur in about 1 in 15,000 births and are associated with distinct neurodevelopmental phenotypes. Prader–Willi syndrome is associated with severe obesity owing to an involuntary urge to eat constantly. Features of Angelman syndrome include severe mental retardation, seizures, and characteristic abnormal behaviors such as a happy, excitable demeanor.

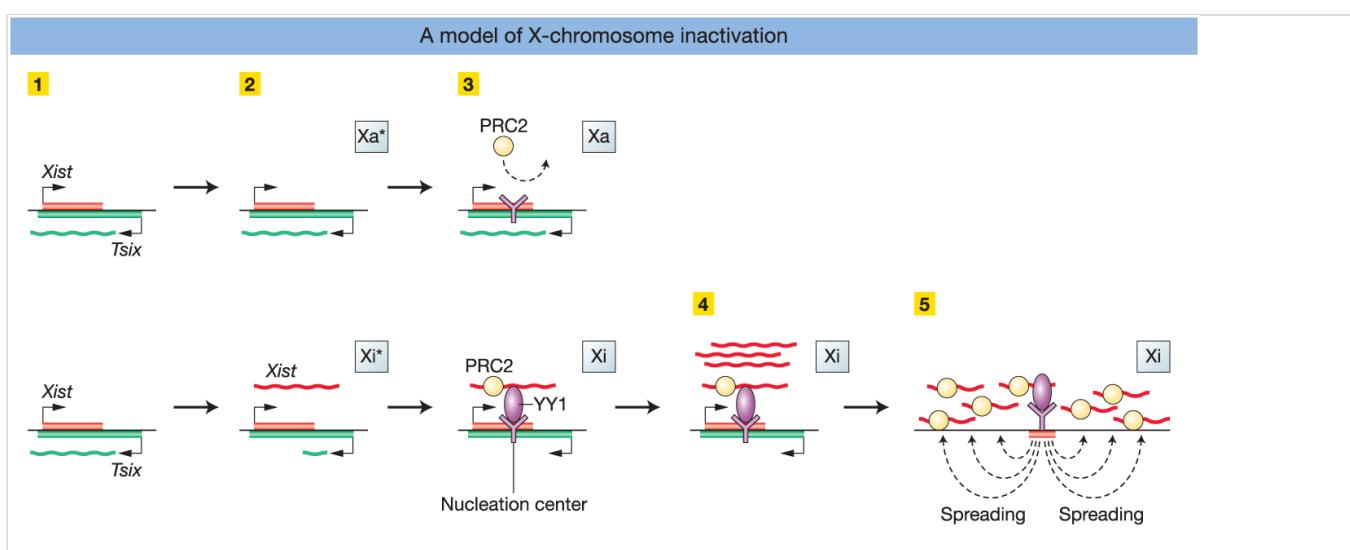
X-chromosome inactivation

Epigenetic regulation of transcription can occur at specific genes, or it can be more global, as in the case of **dosage compensation** in animals. In mammals, females have two X chromosomes and males have only one, creating a potential imbalance in the transcription of genes residing on the X chromosome. This imbalance is corrected by transcriptional silencing of one of the two X chromosomes in females through a process called **X-chromosome inactivation** or X-inactivation, for short. The inactivated X chromosome, called a **Barr body**, can be seen in the nucleus as a darkly staining, highly condensed, heterochromatic structure (**Figure 12-26**). This is a classic example of epigenetic regulation because the two X chromosomes in female cells are nearly identical in sequence; however, one is transcriptionally active and the other is silenced by the formation of heterochromatin. In human cells, the choice of whether to inactivate the maternal or paternal X chromosome is random, but once an X chromosome is inactivated, it will remain inactive for the lifetime of the cell and its daughter cells.



A 17-kilobase-long noncoding RNA (lncRNA) called *Xist* (*X-inactive specific transcript*) plays a central role in initiating silencing of one of the X chromosomes, as does an antisense transcript *Tsix* ("Xist" spelled backward) from the same locus (**Figure 12-27**). Early on in development of the embryo, when both female X chromosomes are transcriptionally active, *Tsix* is expressed from

both alleles. At the beginning of X-chromosome inactivation, transient pairing of the X chromosomes represses the transcription of *Tsix* from one allele, establishing the future inactive X chromosome (Xi). Transcription that persists from the other allele blocks activation of *Xist* transcription and this establishes the future active X chromosome (Xa). *Xist* lncRNA spreads along the future Xi and induces silencing. Spreading is nucleated at the *Xist* locus by the YY1 (Yin-Yang 1) protein, which binds both *Xist* lncRNA and DNA. As it spreads, *Xist* recruits PRC2 (Polycomb Repressive Complex 2), which catalyzes the heterochromatin-associated modification H2K27me3. Other lncRNAs and structural changes in chromatin occur to establish and/or maintain X-chromosome inactivation, including H3K9 methylation, histone deacetylation, DNA CpG island methylation, and incorporation of the histone H2A variant macroH2A into nucleosomes.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 12-27 A model showing how the *Xist* lncRNA, YY1, and PRC2 act to inactivate one X chromosome by forming heterochromatin. The five steps illustrate the progression of events that begin early in embryonic development to silence the transcription of one X chromosome (Xi) and maintain transcription on the other (Xa). Spreading of *Xist* across Xa initiates at a nucleation center at the *Xist* locus. Xi* and Xa* indicate the future inactive and active X chromosomes, respectively.

Dosage compensation also takes place in *Drosophila*, but in this case, the X chromosome is transcriptionally upregulated two-fold in males to equal the transcription of the two X chromosomes in females. As in mammals, dosage compensation in *Drosophila* involves lncRNAs (*roX1* and *roX2*) that are transcribed from the X chromosome in males as well as alterations in chromatin structure that include acetylation of H4K16, phosphorylation of H3S10, and nucleosome remodeling by a chromatin remodeling complex. Twofold transcription upregulation of X-chromosome genes in males is probably achieved by precisely balancing activities that condense and decondense the X chromosome.

KEY CONCEPT In X-inactivation, epigenetic mechanisms enacted early in embryonic development silence an entire chromosome.

SUMMARY

This chapter has focused on the roles that transcription factors and chromatin structure play in directing cell type-specific transcription of eukaryotic genes. Transcription factors bind distal and proximal enhancers and alter transcription by recruiting coactivators and corepressors. A single transcription factor such as Gal4 can affect the transcription of multiple genes by binding an enhancer shared by the genes. In addition, as demonstrated by MCM1 in mating-type specification, one transcription factor can affect the activation/repression activity and transcriptional targets of other transcription factors. Transcription factors are spurred into action by environment signals such as the presence of galactose, hormones, and viruses or developmental signals such as those that specify the intestinal cell fate. Furthermore, cells control transcription by regulating the expression, cellular localization, stability, and activity (DNA binding, dimerization, ligand-binding, and interactions) of transcription factors.

In eukaryotic cells, chromatin, not naked DNA, is the substrate for transcription factors, general transcription factors, and RNA polymerase II. The wrapping of DNA around an octamer of core histones (H2A, H2B, H3, and H4) and the binding of linker histones (H1) to form nucleosomes as well as the gathering of nucleosomes into compact structures is generally repressive to transcription. Reversible chromatin modifying and chromatin remodeling activities work together to overcome the repressive effects of chromatin by changing its structure.

Chromatin modification encompasses modifications to histone amino acids and DNA nucleotides. Modification systems are made up of coactivator and corepressor proteins that add (write), remove (erase), and bind (read) modifications. Protein complexes containing one or several of these activities have affinity for specific genes because of interactions with transcription factors and chromatin modifications themselves. Histone modifications are concentrated in the flexible tails of core histones. There are many different types of modifications, including acetylation, methylation, and phosphorylation, and an enormous number of different combinations of modifications can occur on single histone tails and within a histone octamer. Thus, histone modifications have the capacity to convey complex instructions, dubbed the histone code, for how the transcription machinery should operate. Similarly, DNA modifications, mainly cytosine methylation in CpG dinucleotides, provide an additional set of instructions to the transcription machinery in some eukaryotic organisms, including plants and mammals. Histone and DNA modifications control recruitment of transcription initiation factors such as TFIID and elongation factors such as P-TEFb to specific genes. In addition, they recruit ATP-dependent chromatin remodeling complexes such as SWI/SNF that change factor accessibility to DNA. Chromatin

remodeling complexes make DNA more or less accessible by sliding, ejecting, replacing, or exchanging histones on DNA. Studies such as those of the *IFN- β* gene demonstrate how transcription factors, chromatin modifying factors, and chromatin remodeling factors function cooperatively to adjust the level of transcription of particular genes in response to a signal. In the end, the level of transcription is determined by the balance of activating and repressive mechanisms that act on a gene.

DNA replication faithfully copies both DNA sequence and chromatin structure from parent to daughter cells. Newly formed cells inherit both genetic information, inherent in the nucleotide sequence of DNA, and epigenetic information, built into histone and DNA modifications. Cellular memory, position-effect variegation, genomic imprinting, and X-chromosome inactivation are examples of epigenetic phenomenon where the transcription state of single genes, multiple genes, and even whole chromosomes is inherited without changing the sequence of DNA. Epigenetic inheritance mechanisms involve Polycomb group and Trithorax group factors, suppressor of variegation (*Su(var)*) and enhancer of variegation (*E(var)*) factors, insulator elements, and long noncoding RNAs (lncRNAs). Thus, the nucleotide sequence of genomes is not sufficient for understanding the inheritance of normal and disease states of transcription.

KEY TERMS

[activation domain](#)

[Barr body](#)

[boundary](#)

[canonical histone](#)

[chromatin](#)

[chromatin modification](#)

[chromatin remodeling](#)

[coactivator](#)

[constitutive heterochromatin](#)

[core histone](#)

[core promoter](#)

[corepressor](#)

[CpG island](#)

[dimerization domain](#)

[distal enhancer](#)

[DNA-binding domain](#)

DNA modification
dosage compensation
enhanceosome
enhancer
epigenetic inheritance
euchromatin
facultative heterochromatin
flexible tail
genomic imprinting
general transcription factor (GTF)
heterochromatin
histone
histone acetyltransferase (HAT)
histone code
histone deacetylase (HDAC)
histone fold
histone-fold extension
histone octamer
histone modification
insulator
ligand-binding domain
linker histone
maternal imprinting
nucleosome
nucleosome free region (NFR)
paternal imprinting
position-effect variegation (PEV)
post-translational modification
proximal enhancer
proximal promoter
repression domain
topologically associating domain (TAD)
transcription factor
upstream activation sequence (UAS)
variant histone
X-chromosome inactivation

PROBLEMS

Visit SaplingPlus for supplemental content. Problems with the  icon are available for review/grading.

WORKING WITH THE FIGURES

(The first 27 questions require inspection of text figures.)

1. In [Figure 12-1](#), name the cis-acting sequence elements and trans-acting proteins that regulate transcription. 
2. In [Figure 12-2](#), how might a corepressor block transcription by RNA polymerase II?
3. In [Figure 12-3](#), what proteins bind the GGCCAATC and the ATATAA sequences?
4. In [Figure 12-4](#), what is the function of each transcription factor domain?
5. In [Figure 12-5](#), how does Gal4 regulate the transcription of four different *GAL* genes at the same time?
6. In [Figure 12-6](#), how many individual Gal4 proteins can bind the DNA between the *GAL10* and *GAL1* genes?
7. In [Figure 12-7](#), what effect would a Gal4 protein that lacks the DNA-binding domain have on transcription of the UAS-*lacZ* reporter gene, and why?
8. In [Figure 12-8](#), is Gal3 a transcription factor, coactivator, corepressor, or none of these?
9. In [Figure 12-9](#), hypothesize why MCM1 does not bind and activate α -specific genes in α cells and α/α cells.
10. In [Figure 12-10](#), several protein subunits of the TFIID general transcription factor contain a histone-fold domain. Based on the function of the histone fold in histones, propose a function for the histone fold in TFIID proteins.
11. In [Figure 12-11](#), how might the structure of chromatin bring enhancer and promoter elements close together that are far apart in linear DNA?
12. In [Figure 12-12](#), what features of chromatin structure are shared between enhancers and promoters? 
13. In [Figure 12-13](#), what effect might reduced acetyl-CoA levels have on transcription?

14. Lysines in histone tails can be propionylated. The propionyl group is similar in structure to an acetyl group. Using the categories shown in [Figure 12-14](#), what would you call enzymes that regulate propionyl addition and removal?
15. In [Figure 12-15](#), what are the implications to the histone code of the different lysine and arginine methylation types?
16. In [Figure 12-16](#), how many different codes could be produced on the histone H3 tail just by phosphorylation? 
17. Based on [Figure 12-17](#), what sequence would be read after the sodium bisulfite reaction, if all of the CpGs in the sequence 5'-GGCGTCGAAGTCGAA-3' were methylated? 
18. In [Figure 12-18](#), how might a CpG island function differently than an isolated CpG?
19. In [Figure 12-19](#), what steps would need to occur to exchange a variant H2A for a canonical H2A in a nucleosome? 
20. In [Figure 12-20](#), describe two ways in which the HDAC might be recruited to the *IFN- β* gene.
21. In [Figure 12-21a](#), which of the transcription instructions, in the form of binding sites in chromatin, are reversible?
22. In [Figure 12-22](#), will all flies with a *white* gene inversion have the same pattern of white and red cells as the eye shown at the bottom? Why or why not?
23. In [Figure 12-23](#), name a type of gene that might be an *E(var)* and explain your answer.
24. In [Figure 12-24](#), how is this mechanism similar to the mechanism by which transcription factors regulate transcription (for example, as in [Figure 12-2](#))?
25. In [Figure 12-25](#), what mechanisms might position the enhancer of the paternal allele to act on the *Igf2* promoter that is >50 kilobases away, and why might this not happen for the maternal allele?
26. In [Figure 12-26](#), why is it specified in the figure legend that this is a female cell?
27. In [Figure 12-27](#), what histone modification is expected to be enriched on the inactive X chromosome relative to the active X chromosome, and why?

BASIC PROBLEMS

28. Do all nucleosomes have the same eight core histones? Why or why not?

29. Why might binding of a transcription factor to DNA be inhibited for DNA that is part of a nucleosome?
30. The Luger and Richmond crystal structure of the nucleosome used *Xenopus laevis* (toad) histones. Why is the structure thought to be a good representation of human nucleosomes?
31. Why are histone tails not visible in the crystal structure of the nucleosome core particle?
32. How might higher-order structures of chromatin activate, rather than repress, transcription?
33. What are the two general mechanisms by which histone acetylation affects transcription? 
34. What functions might be served by modifications of amino acids in the histone-fold domain?
35. How is the function of histone tails similar to that of the C-terminal domain (CTD) of RNA polymerase II?
36. By what two mechanisms could histone acetylation levels increase at a gene promoter?
37. Explain how phosphorylation of histone H3 serine 10 (H3S10P) might increase acetylation of histone H4 lysine 16 (H4K16ac)?
38. Vertebrate histone H1 can be phosphorylated on many amino acids in the C-terminal unstructured domain. What effect would you expect histone H1 phosphorylation to have on chromatin structure?
39. What type of factors would you expect to be involved in the regulation of histone H1 phosphorylation?
40. What molecular interactions must be broken by chromatin remodeling complexes to remove a histone octamer from DNA?
41. Why is the order of assembly and disassembly of nucleosomes important for understanding transcription regulation?
42. Why might insertion of a transgene at different places in the *Drosophila* genome cause the transgene to be transcribed at different levels?
43. How would you modify a transgene so that its expression was not affected by position-effect variegation (PEV)?
44. What purpose might be served by the long half-life of core histones?
45. What is meant by the term epigenetic inheritance? Describe two examples of such inheritance. 
46. Give three functions of insulator elements.

47. How many nucleosomes would be needed to cover the human genome (3×10^9 base pairs), if the average linker distance between nucleosomes was 50 base pairs? 
48. Why might the concentration of ATP in cells affect the structure of chromatin?

CHALLENGING PROBLEMS

49. Why is acid used to extract histones from cell nuclei in experiments performed in vitro?
50. A researcher has identified a mutant cell line that has reduced transcription of gene *X* relative to the parental cell line. The mutant cell line has a single point mutation in the entire genome. Describe five possible mechanisms by which the point mutation could reduce the transcription of gene *X*. 
51. For position-effect variegation to have been discovered, why is it critical that the *white* gene is on the X chromosome?
52. How might DNA methylation at a promoter lead to H3K9me3 at nearby nucleosomes?
53. Overexpression of a transcription factor changes the transcription of different genes in different cell types. Why?
54. Draw the pattern of H3K4 trimethylation expected at *IFN-β* promoter during the 24 hours following viral infection ([Figure 12-20a](#)). 
55. Can a transcription factor both activate and repress transcription? Explain your answer.
56. To understand the inheritance of diseases, researchers are mapping genomes and epigenomes (i.e., genomewide chemical modifications to histones and DNA). Describe the information that might be contained in an epigenome map.

GENETICS AND SOCIETY

Accumulating evidence suggests that epigenetic effects can be inherited across multiple generations. For example, the effects of a pregnant woman smoking on her child might also affect their children and their children's children. Do you think that this adds to the moral responsibility of a mother?