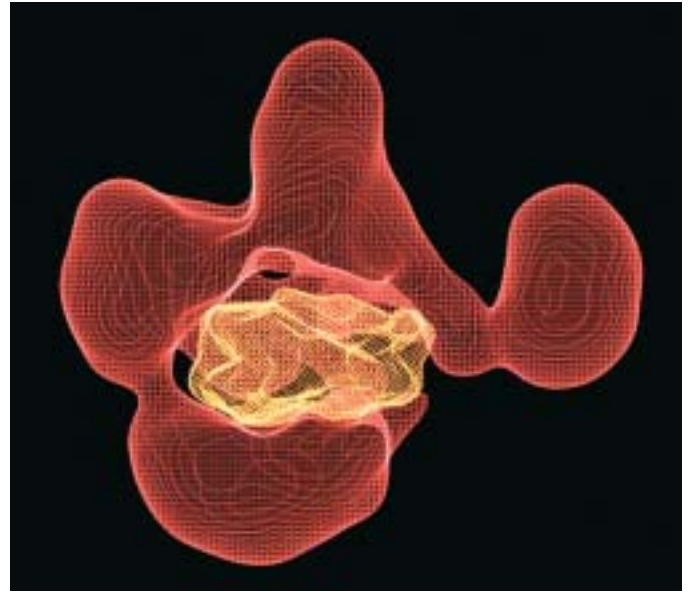


11

TRANSCRIPTIONAL CONTROL OF GENE EXPRESSION



Yeast RSC chromatin remodeling complex structure determined by electron microscopy. The complex (red) is shown bound to a nucleosome (yellow). [Courtesy of Francisco J. Asturias, 2002, *PNAS* **99**:13477]

In previous chapters we've seen that the actions and properties of each cell type are determined by the proteins it contains. In this and the next chapter, we consider how the kinds and amounts of the various proteins produced by a particular cell type in a multicellular organism are regulated. The basic steps in **gene expression**, the entire process whereby the information encoded in a particular gene is decoded into a particular protein, are reviewed in Chapter 4. Synthesis of mRNA requires that an **RNA polymerase** initiate transcription, polymerize ribonucleoside triphosphates complementary to the DNA coding strand, and then terminate transcription (see Figure 4-10). In prokaryotes, ribosomes and translation-initiation factors have immediate access to newly formed RNA transcripts, which function as mRNA without further modification. In eukaryotes, however, the initial RNA transcript is subjected to processing that yields a functional mRNA (see Figure 4-14). The mRNA then is transported from its site of synthesis in the nucleus to the cytoplasm, where it is translated into protein with the aid of ribosomes, tRNAs, and translation factors (see Figure 4-26).

Theoretically, regulation at any one of the various steps in gene expression could lead to *differential* production of proteins in different cell types or developmental stages or in response to external conditions. Although examples of regulation at each step in gene expression have been found, control of transcription initiation—the first step—is the most important mechanism for determining whether most genes are expressed and how much of the encoded mRNAs and, consequently, proteins are produced. The molecular mecha-

nisms that regulate transcription initiation are critical to numerous biological phenomena, including the development of a multicellular organism from a single fertilized egg cell, the immune responses that protect us from pathogenic microorganisms, and neurological processes such as learning and memory. When these regulatory mechanisms function improperly, pathological processes such as cancer and aging may occur.

In this chapter, we focus on the molecular events that determine when transcription of eukaryotic genes is initiated

OUTLINE

- 11.1 Overview of Eukaryotic Gene Control and RNA Polymerases
- 11.2 Regulatory Sequences in Protein-Coding Genes
- 11.3 Activators and Repressors of Transcription
- 11.4 Transcription Initiation by RNA Polymerase II
- 11.5 Molecular Mechanisms of Transcription Activation and Repression
- 11.6 Regulation of Transcription-Factor Activity
- 11.7 Regulated Elongation and Termination of Transcription
- 11.8 Other Eukaryotic Transcription Systems

and also briefly consider control of premature termination of transcription. RNA processing and various posttranscriptional mechanisms for controlling eukaryotic gene expression are covered in the next chapter. Subsequent chapters, particularly Chapters 13, 14, and 15, provide examples of how transcription is regulated by interactions between cells and how the resulting **gene control** contributes to the development and function of specific types of cells in multicellular organisms.

11.1 Overview of Eukaryotic Gene Control and RNA Polymerases

In bacteria, gene control serves mainly to allow a single cell to adjust to changes in its environment so that its growth and division can be optimized. In multicellular organisms, environmental changes also induce changes in gene expression. An example is the response to low oxygen (hypoxia) that is described in Chapter 15. However, the most characteristic and biologically far-reaching purpose of gene control in multicellular organisms is execution of the genetic program that underlies embryological development. Generation of the many different cell types that collectively form a multicellular organism depends on the right genes being activated in the right cells at the right time during the developmental period.

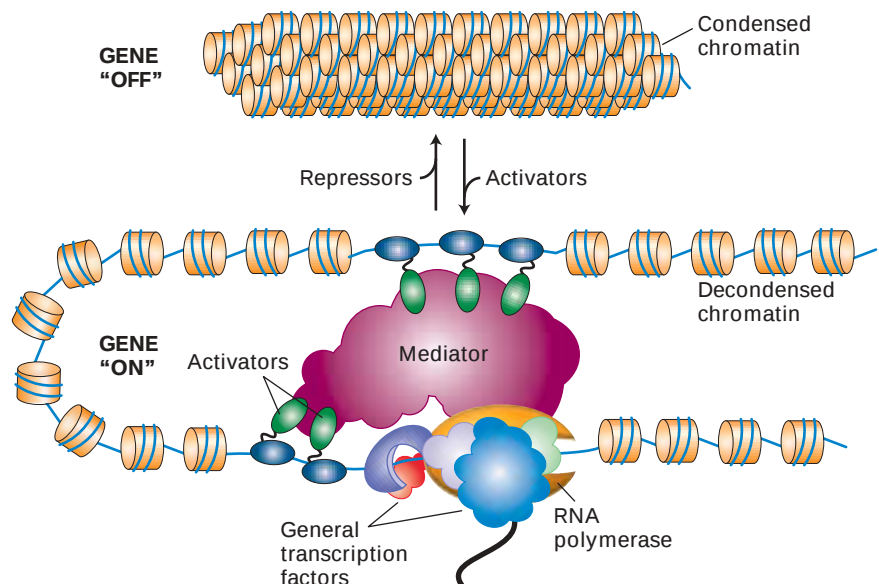
In most cases, once a developmental step has been taken by a cell, it is not reversed. Thus these decisions are fundamentally different from the reversible activation and repression of bacterial genes in response to environmental conditions. In executing their genetic programs, many differentiated cells (e.g., skin cells, red blood cells, and antibody-producing cells) march down a pathway to final cell death, leaving no progeny behind. The fixed patterns of gene con-

trol leading to differentiation serve the needs of the whole organism and not the survival of an individual cell. Despite the differences in the purposes of gene control in bacteria and eukaryotes, two key features of transcription control first discovered in bacteria and described in Chapter 4 also apply to eukaryotic cells. First, protein-binding regulatory DNA sequences, or control elements, are associated with genes. Second, specific proteins that bind to a gene's regulatory sequences determine where transcription will start, and either activate or repress its transcription. As represented in Figure 11-1, in multicellular eukaryotes, inactive genes are assembled into condensed chromatin, which inhibits the binding of RNA polymerases and general transcription factors required for transcription initiation. Activator proteins bind to control elements near the transcription start site of a gene as well as kilobases away and promote chromatin decondensation and binding of RNA polymerase to the promoter. Repressor proteins bind to alternative control elements, causing condensation of chromatin and inhibition of polymerase binding. In this chapter we consider how activators and repressors control chromatin structure and stimulate or inhibit transcription initiation by RNA polymerase.

Most Genes in Higher Eukaryotes Are Regulated by Controlling Their Transcription

Direct measurements of the transcription rates of multiple genes in different cell types have shown that regulation of transcription initiation is the most widespread form of gene control in eukaryotes, as it is in bacteria. *Nascent-chain analysis* is a common method for determining the relative rates of transcription of different genes in cultured cells. In this method, also called *run-on transcription analysis*, isolated nuclei are incubated with ^{32}P -labeled ribonucleoside triphosphates for a brief time (e.g., 5 minutes or less). During

► **FIGURE 11-1 Overview of transcription control in multicellular eukaryotes.** Activator proteins bind to specific DNA control elements in chromatin and interact with multiprotein co-activator machines, such as mediator, to decondense chromatin and assemble RNA polymerase and general transcription factors on promoters. Inactive genes are assembled into regions of condensed chromatin that inhibit RNA polymerases and their associated general transcription factors (GTFs) from interacting with promoters. Alternatively, repressor proteins bind to other control elements to inhibit initiation by RNA polymerase and interact with multiprotein co-repressor complexes to condense chromatin.



this incubation period, RNA polymerase molecules that were actively transcribing genes when the nuclei were isolated incorporate several hundred labeled nucleotides into each nascent (growing) RNA chain, but little initiation of new chains occurs. The total radioactive label incorporated into RNA is a measure of the overall transcription rate. The fraction of the total labeled RNA produced by transcription of a particular gene—that is, its relative transcription rate—is determined by hybridizing the labeled RNA to the cloned DNA of that gene attached to a membrane.

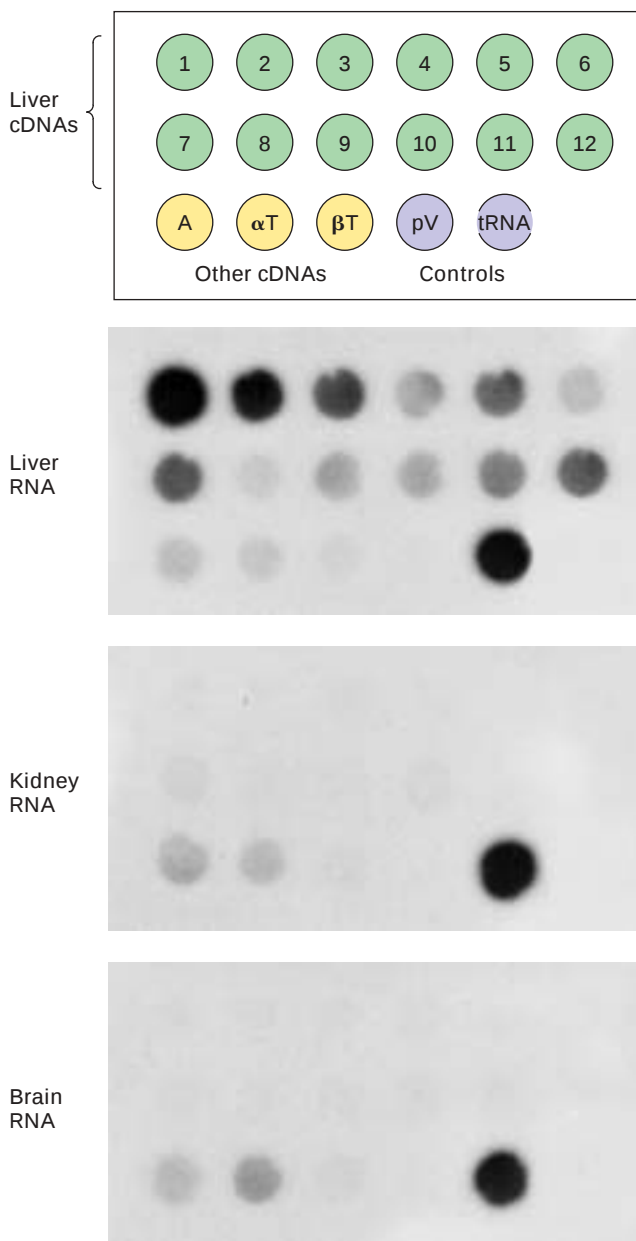
The results of run-on transcription analyses illustrated in Figure 11-2 show that transcription of genes encoding proteins expressed specifically in hepatocytes (the major cell type in mammalian liver) is readily detected in nuclei pre-

pared from liver, but not in nuclei from brain or kidney. Since the run-on transcription assay measures RNA synthesis, these results indicate that differential synthesis of liver-specific proteins is regulated by controlling transcription of the corresponding genes in different tissues. Similar results have been obtained in run-on transcription experiments with other cell types and a wide variety of tissue-specific proteins, indicating that transcriptional control is the primary mechanism of gene control in complex organisms.

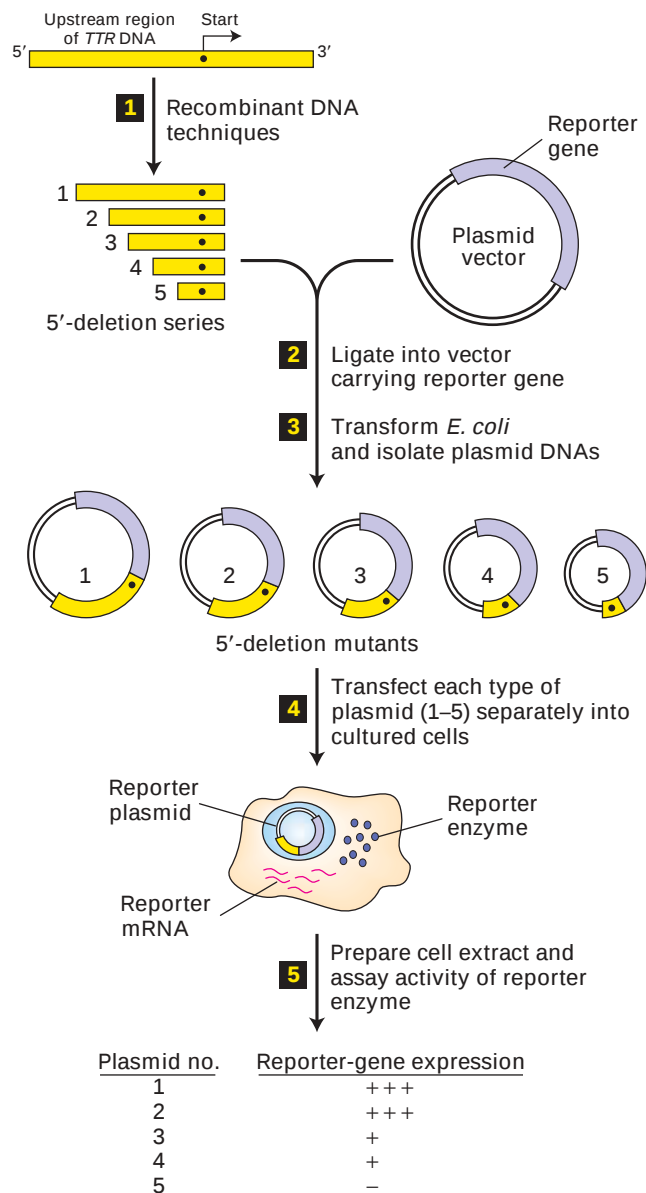
Regulatory Elements in Eukaryotic DNA Often Are Many Kilobases from Start Sites

In eukaryotes, as in bacteria, a DNA sequence that specifies where RNA polymerase binds and initiates transcription of a gene is called a **promoter**. Transcription from a particular promoter is controlled by DNA-binding proteins, termed **transcription factors**, that are equivalent to bacterial *repressors* and *activators*. However, the DNA control elements in eukaryotic genomes that bind transcription factors often are located much farther from the promoter they regulate than is the case in prokaryotic genomes. In some cases, transcription factors that regulate expression of protein-coding genes in higher eukaryotes bind at regulatory sites tens of thousands of base pairs either **upstream** (opposite to the direction of transcription) or **downstream** (in the same direction as transcription) from the promoter. As a result of this arrangement, transcription from a single promoter may be regulated by binding of multiple transcription factors to alternative control elements, permitting complex control of gene expression.

For example, alternative transcription-control elements regulate expression of the mammalian gene that encodes transthyretin (*TTR*), which transports thyroid hormone in blood and the cerebrospinal fluid that surrounds the brain



◀ **EXPERIMENTAL FIGURE 11-2 Measurement of RNA production in various tissues demonstrates that transcription of a given gene generally occurs only in the cell types in which it is expressed.** Nuclei from mouse liver, kidney, and brain cells were exposed to ^{32}P -UTP. The three resulting labeled RNA samples were hybridized to separate nitrocellulose membranes; on each membrane was fixed an identical pattern of various cDNAs (*top*). The cDNAs labeled 1–12 (green) encode proteins synthesized actively in liver (e.g., albumin, transferrin) but not in most other tissues. The other cDNAs tested were actin (A) and α - and β -tubulin (αT , βT) (yellow), which are proteins found in almost all cell types. A gene encoding methionine tRNA and the plasmid vector DNA (pV) in which the cDNAs were cloned were included as controls (purple). After removal of unhybridized RNAs, the labeled RNAs complementary to the cDNAs were revealed by autoradiography. RNA from the liver sample hybridized extensively with the cDNAs. Little hybridization is seen for the kidney and brain samples, indicating that genes encoding proteins found in hepatocytes are transcribed in liver cells but not in the cells of the other tissues. [See E. Derman et al., 1981, *Cell* **23**:731, and D. J. Powell et al., 1984, *J. Mol. Biol.* **197**:21]



▲ **EXPERIMENTAL FIGURE 11-3 5'-Deletion analysis can identify transcription-control sequences in DNA upstream of a eukaryotic gene.** Step **1**: Recombinant DNA techniques are used to prepare a series of DNA fragments that extend from the 5'-untranslated region of a gene various distances upstream. Step **2**: The DNA fragments are ligated into a reporter plasmid upstream of an easily assayed reporter gene. Step **3**: The DNA is transformed into *E. coli* to isolate plasmids with deletions of various lengths 5' to the transcription start site. Step **4**: Each plasmid is then transfected into cultured cells (or used to prepare transgenic organisms) and expression of the reporter gene is assayed (step **5**). The results of this hypothetical example (bottom) indicate that the test fragment contains two control elements. The 5' end of one lies between deletions 2 and 3; the 5' end of the other lies between deletions 4 and 5.

and spinal cord. Transthyretin is expressed in hepatocytes, which synthesize and secrete most of the blood serum proteins, and in choroid plexus cells in the brain, which secrete cerebrospinal fluid and its constituent proteins. The control elements required for transcription of the *TTR* gene were identified by the procedure outlined in Figure 11-3. In this experimental approach, DNA fragments with varying extents of sequence upstream of a start site are cloned in front of a **reporter gene** in a bacterial plasmid using recombinant DNA techniques. Reporter genes express enzymes that are easily assayed in cell extracts. Commonly used reporter genes include the *E. coli lacZ* gene encoding β -galactosidase; the firefly gene encoding luciferase, which converts energy from ATP hydrolysis into light; and the jellyfish gene encoding green fluorescent protein (GFP).

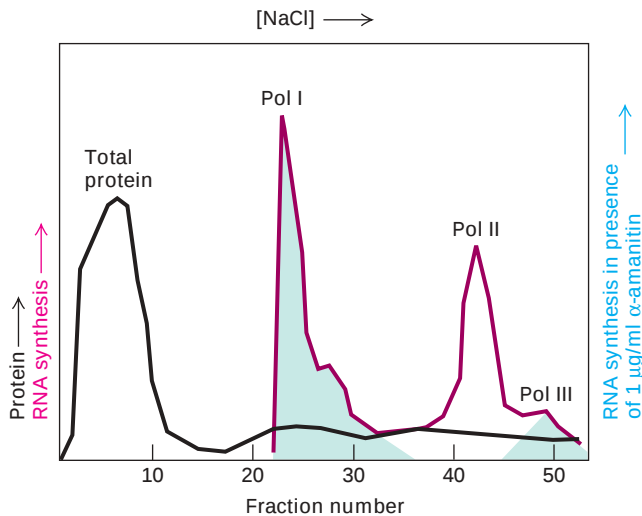
By constructing and analyzing a 5'-deletion series upstream of the *TTR* gene, researchers identified two control elements that stimulate reporter-gene expression in hepatocytes, but not in other cell types. One region mapped between ≈ 2.01 and 1.85 kb upstream of the *TTR* gene start site; the other mapped between ≈ 200 base pairs upstream and the start site. Further studies demonstrated that alternative DNA sequences control *TTR* transcription in choroid plexus cells. Thus, alternative control elements regulate transcription of the *TTR* gene in two different cell types. We examine the basic molecular events underlying this type of eukaryotic transcriptional control in later sections.

Three Eukaryotic Polymerases Catalyze Formation of Different RNAs

The nuclei of all eukaryotic cells examined so far (e.g., vertebrate, *Drosophila*, yeast, and plant cells) contain three different RNA polymerases, designated I, II, and III. These enzymes are eluted at different salt concentrations during ion-exchange chromatography and also differ in their sensitivity to α -amanitin, a poisonous cyclic octapeptide produced by some mushrooms (Figure 11-4). Polymerase I is very insensitive to α -amanitin; polymerase II is very sensitive; and polymerase III has intermediate sensitivity.

Each eukaryotic RNA polymerase catalyzes transcription of genes encoding different classes of RNA. *RNA polymerase I*, located in the nucleolus, transcribes genes encoding precursor rRNA (**pre-rRNA**), which is processed into 28S, 5.8S, and 18S rRNAs. *RNA polymerase III* transcribes genes encoding tRNAs, 5S rRNA, and an array of small, stable RNAs, including one involved in RNA splicing (U6) and the RNA component of the signal-recognition particle (SRP) involved in directing nascent proteins to the endoplasmic reticulum (Chapter 16). *RNA polymerase II* transcribes all protein-coding genes; that is, it functions in production of mRNAs. RNA polymerase II also produces four of the five small nuclear RNAs that take part in RNA splicing.

Each of the three eukaryotic RNA polymerases is more complex than *E. coli* RNA polymerase, although their structures are similar (Figure 11-5). All three contain two large



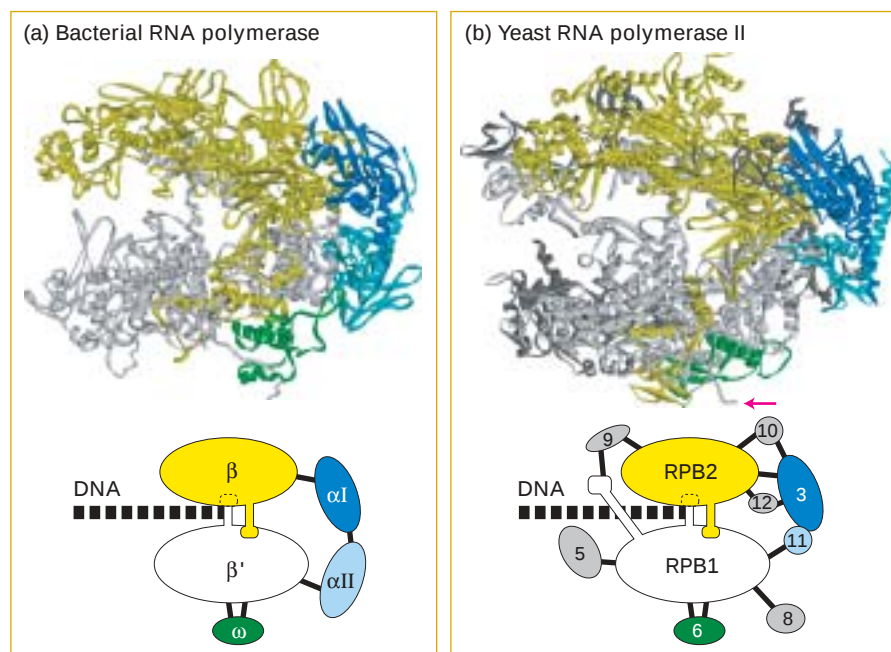
◀ **EXPERIMENTAL FIGURE 11-4** Column chromatography separates and identifies the three eukaryotic RNA polymerases, each with its own sensitivity to α -amanitin.

A protein extract from the nuclei of cultured eukaryotic cells is passed through a DEAE Sephadex column and adsorbed protein eluted (black curve) with a solution of constantly increasing NaCl concentration. Three fractions from the eluate subsequently showed RNA polymerase activity (red curve). At a concentration of 1 $\mu\text{g/ml}$, α -amanitin inhibits polymerase II activity but has no effect on polymerases I and III (green shading). Polymerase III is inhibited by 10 $\mu\text{g/ml}$ of α -amanitin, whereas polymerase I is unaffected even at this higher concentration. [See R. G. Roeder, 1974, *J. Biol. Chem.* **249**:241]

subunits and 10–14 smaller subunits, some of which are present in two or all three of the polymerases. The best-characterized eukaryotic RNA polymerases are from the yeast *S. cerevisiae*. Each of the yeast genes encoding the polymerase subunits has been cloned and sequenced and the effects of gene-knockout mutations have been characterized. In addition, the three-dimensional structure of yeast RNA

polymerase II missing two nonessential subunits has been determined (see Figure 11-5). The three nuclear RNA polymerases from all eukaryotes so far examined are very similar to those of yeast.

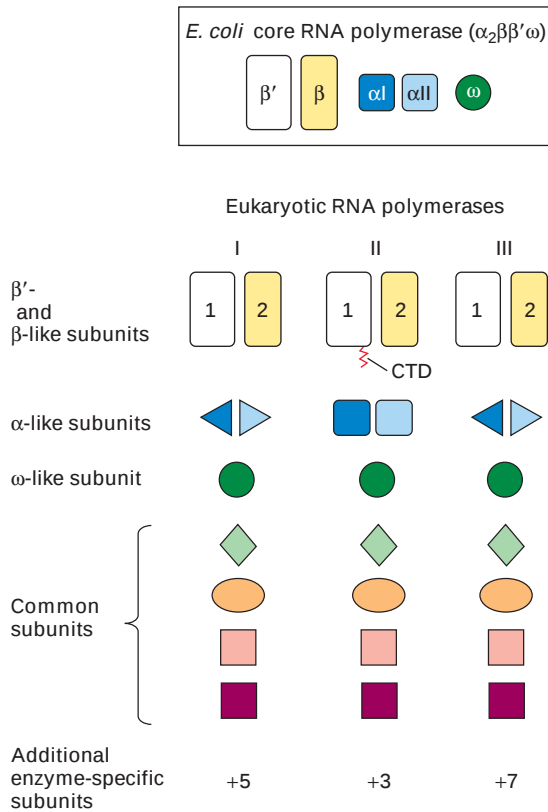
The two large subunits (RPB1 and RPB2) of all three eukaryotic RNA polymerases are related to each other and are similar to the *E. coli* β' and β subunits, respectively



▲ **FIGURE 11-5** Comparison of three-dimensional structures of bacterial and eukaryotic RNA polymerases.

These C_α trace models are based on x-ray crystallographic analysis of RNA polymerase from the bacterium *T. aquaticus* and RNA polymerase II from *S. cerevisiae*. (a) The five subunits of the bacterial enzyme are distinguished by color. Only the N-terminal domains of the α subunits are included in this model. (b) Ten of the twelve subunits constituting yeast RNA polymerase II are shown in this model.

Subunits that are similar in conformation to those in the bacterial enzyme are shown in the same colors. The C-terminal domain of the large subunit RPB1 was not observed in the crystal structure, but it is known to extend from the position marked with a red arrow. (RPB is the abbreviation for “RNA polymerase B,” which is an alternative way of referring to RNA polymerase II.) [Part (a) based on crystal structures from G. Zhang et al., 1999, *Cell* **98**:811. Part (b) from P. Cramer et al., 2001, *Science* **292**:1863.]



▲ **FIGURE 11-6 Schematic representation of the subunit structure of the *E. coli* RNA core polymerase and yeast nuclear RNA polymerases.** All three yeast polymerases have five core subunits homologous to the β , β' , two α , and ω subunits of *E. coli* RNA polymerase. The largest subunit (RBP1) of RNA polymerase II also contains an essential C-terminal domain (CTD). RNA polymerases I and III contain the same two nonidentical α -like subunits, whereas RNA polymerase II contains two other nonidentical α -like subunits. All three polymerases share the same ω -like subunit and four other common subunits. In addition, each yeast polymerase contains three to seven unique smaller subunits.

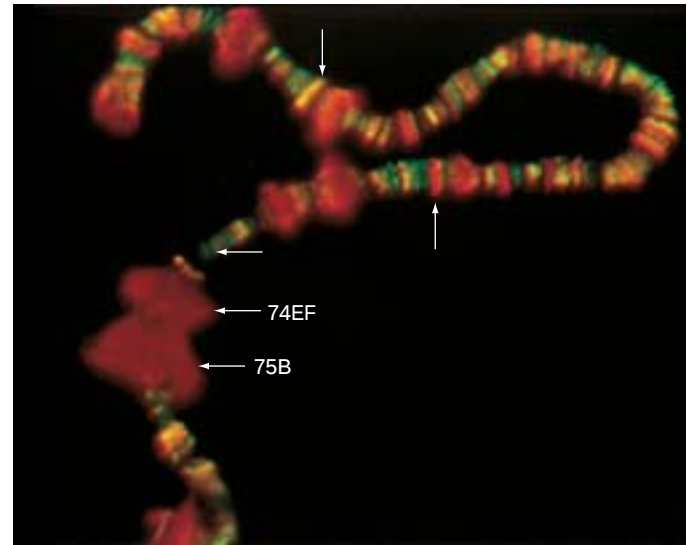
(Figure 11-6). Each of the eukaryotic polymerases also contains an ω -like and two nonidentical α -like subunits. The extensive similarity in the structures of these core subunits in RNA polymerases from various sources indicates that this enzyme arose early in evolution and was largely conserved. This seems logical for an enzyme catalyzing a process so basic as copying RNA from DNA.

In addition to their core subunits related to the *E. coli* RNA polymerase subunits, all three yeast RNA polymerases contain four additional small subunits, common to them but not to the bacterial RNA polymerase. Finally, each eukaryotic polymerase has several enzyme-specific subunits that are not present in the other two polymerases. Gene-knockout experiments in yeast indicate that most of these subunits are essential for cell viability. Disruption of the few polymerase subunit genes that are not absolutely essential for viability nevertheless results in very poorly growing cells. Thus it

seems likely that all the subunits are necessary for eukaryotic RNA polymerases to function normally.

The Largest Subunit in RNA Polymerase II Has an Essential Carboxyl-Terminal Repeat

The carboxyl end of the largest subunit of RNA polymerase II (RBP1) contains a stretch of seven amino acids that is nearly precisely repeated multiple times. Neither RNA polymerase I nor III contains these repeating units. This heptapeptide repeat, with a consensus sequence of Tyr-Ser-Pro-Thr-Ser-Pro-Ser, is known as the *carboxyl-terminal domain* (CTD). Yeast RNA polymerase II contains 26 or more repeats, the mammalian enzyme has 52 repeats, and an intermediate number of repeats occur in RNA polymerase II from nearly all other eukaryotes. The CTD is critical for viability, and at least 10 copies of the repeat must be present for yeast to survive.



▲ **EXPERIMENTAL FIGURE 11-7 Antibody staining demonstrates that the carboxyl-terminal domain (CTD) of RNA polymerase II is phosphorylated during *in vivo* transcription.** Salivary gland polytene chromosomes were prepared from *Drosophila* larvae just before molting. The preparation was treated with a rabbit antibody specific for phosphorylated CTD and with a goat antibody specific for unphosphorylated CTD. The preparation then was stained with fluorescein-labeled anti-goat antibody (green) and rhodamine-labeled anti-rabbit antibody (red). Thus polymerase molecules with an unphosphorylated CTD stain green, and those with a phosphorylated CTD stain red. The molting hormone ecdysone induces very high rates of transcription in the puffed regions labeled 74EF and 75B; note that only phosphorylated CTD is present in these regions. Smaller puffed regions transcribed at high rates also are visible. Nonpuffed sites that stain red (up arrow) or green (horizontal arrow) also are indicated, as is a site staining both red and green, producing a yellow color (down arrow). [From J. R. Weeks et al., 1993, *Genes & Dev.* 7:2329; courtesy of J. R. Weeks and A. L. Greenleaf.]

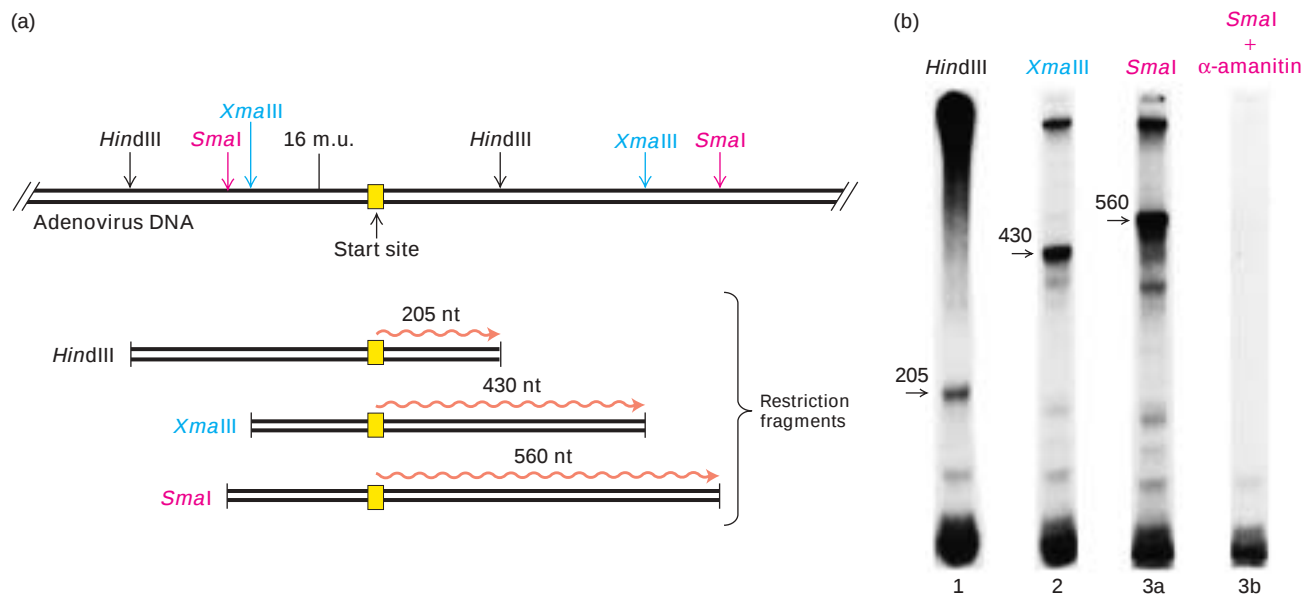
In vitro experiments with model promoters first showed that RNA polymerase II molecules that initiate transcription have an unphosphorylated CTD. Once the polymerase initiates transcription and begins to move away from the promoter, many of the serine and some tyrosine residues in the CTD are phosphorylated. Analysis of polytene chromosomes from *Drosophila* salivary glands prepared just before molting of the larva indicate that the CTD also is phosphorylated during in vivo transcription. The large chromosomal “puffs” induced at this time in development are regions where the genome is very actively transcribed. Staining with antibodies specific for the phosphorylated or unphosphorylated CTD demonstrated that RNA polymerase II associated with the highly transcribed puffed regions contains a phosphorylated CTD (Figure 11-7).

RNA Polymerase II Initiates Transcription at DNA Sequences Corresponding to the 5' Cap of mRNAs

Several experimental approaches have been used to identify DNA sequences at which RNA polymerase II initiates transcription. Approximate mapping of the transcription start

site is possible by exposing cultured cells or isolated nuclei to ^{32}P -labeled ribonucleotides for very brief times, as described earlier. After the resulting labeled nascent transcripts are separated on the basis of chain length, each size fraction is incubated under hybridization conditions with overlapping restriction fragments that encompass the DNA region of interest. The restriction fragment that hybridizes to the shortest labeled nascent chain, as well as all the longer ones, contains the transcription start site. One of the first transcription units to be analyzed in this way was the major late transcription unit of adenovirus. This analysis indicated that transcription was initiated in a region ≈ 6 kb from the left end of the viral genome.

The precise base pair where RNA polymerase II initiates transcription in the adenovirus late transcription unit was determined by analyzing the RNAs synthesized during in vitro transcription of adenovirus DNA restriction fragments that extended somewhat upstream and downstream of the approximate initiation region determined by nascent-transcript analysis. The rationale of this experiment and typical results are illustrated in Figure 11-8. The RNA transcripts synthesized in vitro by RNA polymerase II from the



▲ EXPERIMENTAL FIGURE 11-8 In vitro transcription of restriction fragments and measurement of the RNA lengths localize the initiation site of the adenovirus major late transcription unit. (a) The top line shows restriction sites for *Hind*III (black), *Xma*III (blue), and *Sma*I (red) in the region of the adenovirus genome where the transcription-initiation site was located by nascent-transcript analysis (near 16 map units). The *Hind*III, *Xma*III, and *Sma*I restriction fragments that encompass the initiation site were individually incubated with a nuclear extract prepared from cultured cells and ^{32}P -labeled ribonucleoside triphosphates. Transcription of each fragment begins at the start site and ends when an RNA polymerase II molecule “runs off” the cut end of the fragment template, producing a run-off transcript (wavy red lines). (b) The run-off transcripts synthesized

from each fragment were then subjected to gel electrophoresis and autoradiography to determine their exact lengths. Since the positions of the restriction sites in the adenovirus DNA were known, the lengths of the run-off transcripts in nucleotides (nt) produced from the restriction fragments precisely map the initiation site on the adenovirus genome, as diagrammed in part (a). In the gels shown here, the bands at the top and bottom represent high- and low-molecular-weight RNA transcripts that are formed under the conditions of the experiment. The sample in lane 3b is the same as that in lane 3a, except that α -amanitin, an inhibitor of RNA polymerase II, was included in the transcription mixture. See the text for further discussion. [See R. M. Evans and E. Ziff, 1978, *Cell* **15**:1463, and P. A. Weil et al., 1979, *Cell* **18**:469. Autoradiogram courtesy of R. G. Roeder.]

start site determined in this experiment contained an RNA cap structure identical with that present at the 5' end of nearly all eukaryotic mRNAs (see Figure 4-13). This 5' cap was added by enzymes in the nuclear extract, which can add a cap only to an RNA that has a 5' tri- or diphosphate. Because a 5' end generated by cleavage of a longer RNA would have a 5' monophosphate, it could not be capped. Consequently, researchers concluded that the capped nucleotide generated in the *in vitro* transcription reaction must have been the nucleotide with which transcription was initiated. The finding that the sequence at the 5' end of the RNA transcripts produced *in vitro* is the same as that at the 5' end of late adenovirus mRNAs isolated from cells confirmed that the capped nucleotide of adenovirus late mRNAs coincides with the transcription-initiation site.

Similar *in vitro* transcription assays with other cloned eukaryotic genes have produced similar results. In each case, the start site was found to be equivalent to the capped 5' sequence of the corresponding mRNA. Thus synthesis of eukaryotic precursors of mRNAs by RNA polymerase II begins at the DNA sequence encoding the capped 5' end of the mRNA. Today, the transcription start site for a newly characterized mRNA generally is determined simply by identifying the DNA sequence encoding the 5' end of the mRNA.

KEY CONCEPTS OF SECTION 11.1

Overview of Eukaryotic Gene Control and RNA Polymerases

- The primary purpose of gene control in multicellular organisms is the execution of precise developmental decisions so that the proper genes are expressed in the proper cells during development and cellular differentiation.
- Transcriptional control is the primary means of regulating gene expression in eukaryotes, as it is in bacteria.
- In eukaryotic genomes, DNA transcription control elements may be located many kilobases away from the promoter they regulate. Different control regions can control transcription of the same gene in different cell types.
- Eukaryotes contain three types of nuclear RNA polymerases. All three contain two large and three smaller core subunits with homology to the β' , β , α , and ω subunits of *E. coli* RNA polymerase, as well several additional small subunits (see Figure 11-6).
- RNA polymerase I synthesizes only pre-rRNA. RNA polymerase II synthesizes mRNAs and some of the small nuclear RNAs that participate in mRNA splicing. RNA polymerase III synthesizes tRNAs, 5S rRNA, and several other relatively short, stable RNAs.
- The carboxyl-terminal domain (CTD) in the largest subunit of RNA polymerase II becomes phosphorylated during

transcription initiation and remains phosphorylated as the enzyme transcribes the template.

- RNA polymerase II initiates transcription of genes at the nucleotide in the DNA template that corresponds to the 5' nucleotide that is capped in the encoded mRNA.

11.2 Regulatory Sequences in Protein-Coding Genes

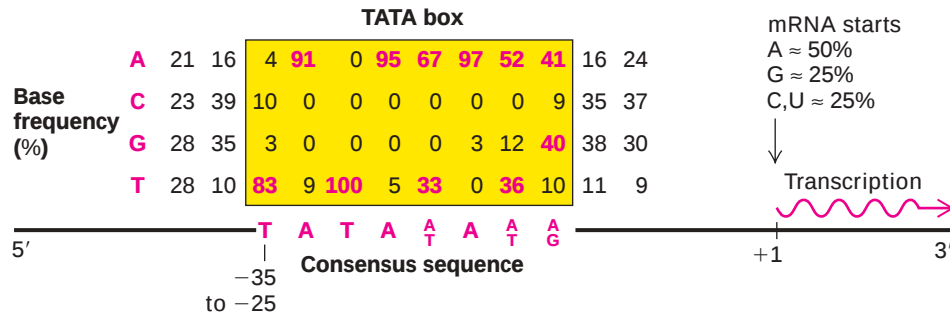
As noted in the previous section, expression of eukaryotic protein-coding genes is regulated by multiple protein-binding DNA sequences, generically referred to as **transcription-control regions**. These include promoters and other types of control elements located near transcription start sites, as well as sequences located far from the genes they regulate. In this section, we take a closer look at the properties of various control elements found in eukaryotic protein-coding genes and some techniques used to identify them.

The TATA Box, Initiators, and CpG Islands Function as Promoters in Eukaryotic DNA

The first genes to be sequenced and studied in *in vitro* transcription systems were viral genes and cellular protein-coding genes that are very actively transcribed either at particular times of the cell cycle or in specific differentiated cell types. In all these rapidly transcribed genes, a conserved sequence called the **TATA box** was found ≈ 25 – 35 base pairs upstream of the start site (Figure 11-9). Mutagenesis studies have shown that a single-base change in this nucleotide sequence drastically decreases *in vitro* transcription by RNA polymerase II of genes adjacent to a TATA box. In most cases, sequence changes between the TATA box and start site do not significantly affect the transcription rate. If the base pairs between the TATA box and the normal start site are deleted, transcription of the altered, shortened template begins at a new site ≈ 25 base pairs downstream from the TATA box. Consequently, the TATA box acts similarly to an *E. coli* promoter to position RNA polymerase II for transcription initiation (see Figure 4-11).

Instead of a TATA box, some eukaryotic genes contain an alternative promoter element called an *initiator*. Most naturally occurring initiator elements have a cytosine (C) at the -1 position and an adenine (A) residue at the transcription start site ($+1$). Directed mutagenesis of mammalian genes with an initiator-containing promoter has revealed that the nucleotide sequence immediately surrounding the start site determines the strength of such promoters. Unlike the conserved TATA box sequence, however, only an extremely degenerate initiator consensus sequence has been defined:





▲ **FIGURE 11-9 Determination of consensus TATA box sequence.** The nucleotide sequences upstream of the start site in 900 different eukaryotic protein-coding genes were aligned to maximize homology in the region from -35 to -26 . The tabulated numbers are the percentage frequency of each base at each position. Maximum homology occurs over an eight-base

region, referred to as the *TATA box*, whose consensus sequence is shown at the bottom. The initial base in mRNAs encoded by genes containing a TATA box most frequently is an A. [See P. Bucher, 1990, *J. Mol. Biol.* **212**:563, and http://www.epd.isb-sib.ch/promoter_elements.]

where A^{+1} is the base at which transcription starts, Y is a pyrimidine (C or T), N is any of the four bases, and T/A is T or A at position $+3$.

Transcription of genes with promoters containing a TATA box or initiator element begins at a well-defined initiation site. However, transcription of many protein-coding genes has been shown to begin at any one of multiple possible sites over an extended region, often 20–200 base pairs in length. As a result, such genes give rise to mRNAs with multiple alternative 5' ends. These genes, which generally are transcribed at low rates (e.g., genes encoding the enzymes of intermediary metabolism, often called “housekeeping genes”), do not contain a TATA box or an initiator. Most genes of this type contain a CG-rich stretch of 20–50 nucleotides within ≈ 100 base pairs upstream of the start-site region. The dinucleotide CG is statistically underrepresented in vertebrate DNAs, and the presence of a CG-rich region, or *CpG island*, just upstream from a start site is a distinctly nonrandom distribution. For this reason, the presence of a CpG island in genomic DNA suggests that it may contain a transcription-initiation region.

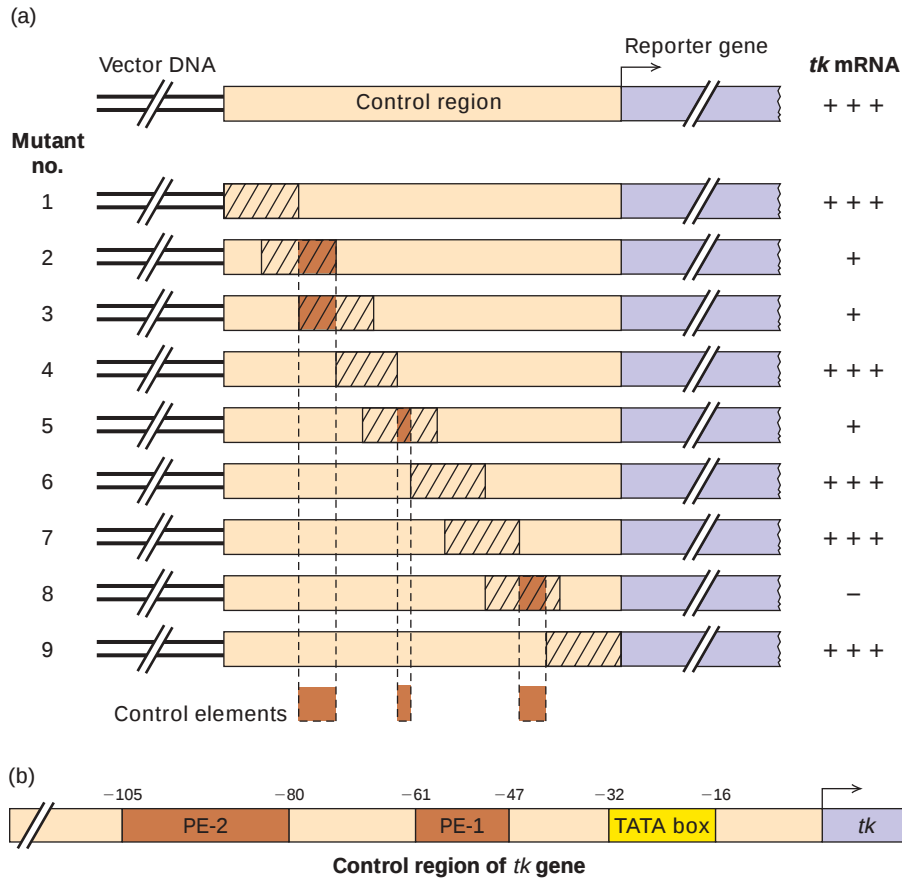
Promoter-Proximal Elements Help Regulate Eukaryotic Genes

Recombinant DNA techniques have been used to systematically mutate the nucleotide sequences upstream of the start sites of various eukaryotic genes in order to identify transcription-control regions. By now, hundreds of eukaryotic genes have been analyzed, and scores of transcription-control regions have been identified. These control elements, together with the TATA-box or initiator, often are referred to as the *promoter* of the gene they regulate. However, we prefer to reserve the term *promoter* for the TATA-box or initiator sequences that determine the initiation site in the template. We use the term **promoter-proximal elements** for

control regions lying within 100–200 base pairs upstream of the start site. In some cases, promoter-proximal elements are cell-type-specific; that is, they function only in specific differentiated cell types.

One approach frequently taken to determine the upstream border of a transcription-control region for a mammalian gene involves constructing a set of 5' deletions as discussed earlier (see Figure 11-3). Once the 5' border of a transcription-control region is determined, analysis of *linker scanning mutations* can pinpoint the sequences with regulatory functions that lie between the border and the transcription start site. In this approach, a set of constructs with contiguous overlapping mutations are assayed for their effect on expression of a reporter gene or production of a specific mRNA (Figure 11-10a). One of the first uses of this type of analysis identified promoter-proximal elements of the thymidine kinase (*tk*) gene from herpes simplex virus (HSV). The results demonstrated that the DNA region upstream of the HSV *tk* gene contains three separate transcription-control sequences: a TATA box in the interval from -32 to -16 , and two other control elements farther upstream (Figure 11-10b).

To test the spacing constraints on control elements in the HSV *tk* promoter region identified by analysis of linker scanning mutations, researchers prepared and assayed constructs containing small deletions and insertions between the elements. Changes in spacing between the promoter and promoter-proximal control elements of 20 nucleotides or fewer had little effect. However, insertions of 30 to 50 base pairs between a promoter-proximal element and the TATA box was equivalent to deleting the element. Similar analyses of other eukaryotic promoters have also indicated that considerable flexibility in the spacing between promoter-proximal elements is generally tolerated, but separations of several tens of base pairs may decrease transcription.



▲ **EXPERIMENTAL FIGURE 11-10 Linker scanning mutations identify transcription-control elements.** (a) A region of eukaryotic DNA (orange) that supports high-level expression of a reporter gene (light blue) is cloned in a plasmid vector as diagrammed at the top. Overlapping linker scanning (LS) mutations (crosshatch) are introduced from one end of the region being analyzed to the other. These mutations result from scrambling the nucleotide sequence in a short stretch of the DNA. After the mutant plasmids are transfected separately into cultured cells, the activity of the reporter-gene product is assayed. In the hypothetical example shown here, LS mutations

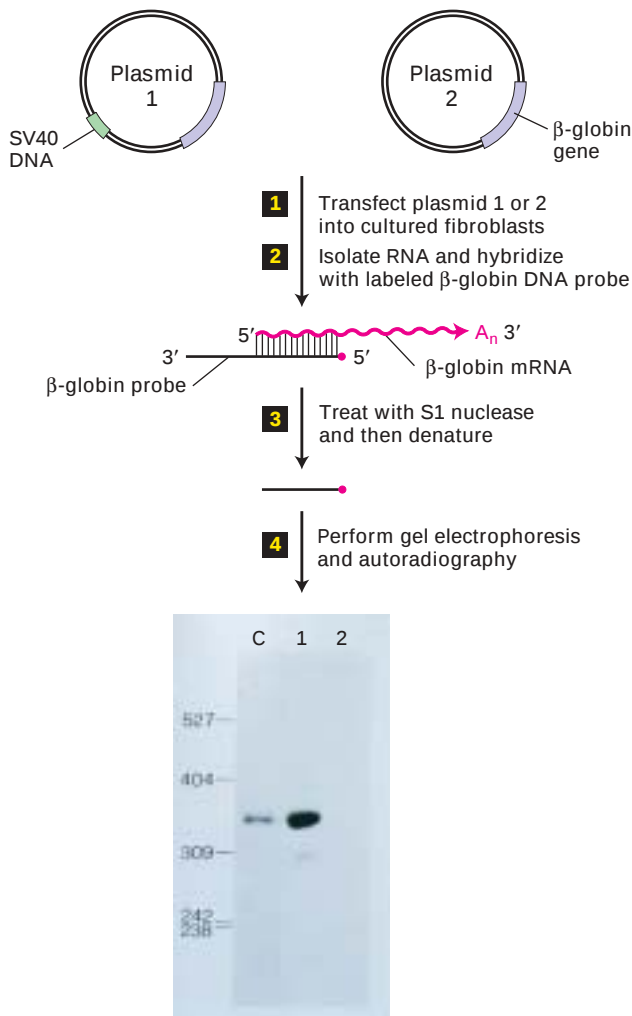
1, 4, 6, 7, and 9 have little or no effect on expression of the reporter gene, indicating that the regions altered in these mutants contain no control elements. Reporter-gene expression is significantly reduced in mutants 2, 3, 5, and 8, indicating that control elements (brown) lie in the intervals shown at the bottom. (b) Analysis of LS mutations in the transcription-control region of the thymidine kinase (*tk*) gene from herpes simplex virus (HSV) identified a TATA box and two promoter-proximal elements (PE-1 and PE-2). [Part (b) see S. L. McKnight and R. Kingsbury, 1982, *Science* **217**:316.]

Distant Enhancers Often Stimulate Transcription by RNA Polymerase II

As noted earlier, transcription from many eukaryotic promoters can be stimulated by control elements located thousands of base pairs away from the start site. Such long-distance transcription-control elements, referred to as **enhancers**, are common in eukaryotic genomes but fairly rare in bacterial genomes. The first enhancer to be discovered that stimulates transcription of eukaryotic genes was in a 366-bp fragment of the simian virus 40 (SV40) genome (Figure 11-11). Further analysis of this region of SV40 DNA revealed that an ≈100-bp sequence lying ≈100 base pairs upstream of the SV40 early transcription start site was

responsible for its ability to enhance transcription. In SV40, this enhancer sequence functions to stimulate transcription from viral promoters. The SV40 enhancer, however, stimulates transcription from all mammalian promoters that have been tested when it is inserted in either orientation anywhere on a plasmid carrying the test promoter, even when it is thousands of base pairs from the start site. An extensive linker scanning mutational analysis of the SV40 enhancer indicated that it is composed of multiple individual elements, each of which contributes to the total activity of the enhancer. As discussed later, each of these regulatory elements is a protein-binding site.

Soon after discovery of the SV40 enhancer, enhancers were identified in other viral genomes and in eukaryotic



▲ EXPERIMENTAL FIGURE 11-11 Plasmids containing a particular SV40 DNA fragment showed marked increase in mRNA production compared with plasmids lacking this enhancer. Plasmids containing the β -globin gene with or without a 366-bp fragment of SV40 DNA were constructed. These plasmids were transfected into cultured cells, and any resulting RNA was hybridized to a β -globin DNA probe (steps **1** and **2**). The amount of β -globin mRNA synthesized by cells transfected with one or the other plasmid was assayed by the S1 nuclease-protection method (step **3**). The restriction-fragment probe, generated from a β -globin cDNA clone, was complementary to the 5' end of β -globin mRNA. The 5' end of the probe was labeled with ^{32}P (red dot). Hybridization of β -globin mRNA to the probe protected an ≈ 340 -nucleotide fragment of the probe from digestion by S1 nuclease, which digests single-stranded DNA but not DNA in an RNA-DNA hybrid. Autoradiography of electrophoresed S1-protected fragments (step **4**) revealed that cells transfected with plasmid 1 (lane 1) produced much more β -globin mRNA than those transfected with plasmid 2 (lane 2). Lane C is a control assay of β -globin mRNA isolated from reticulocytes, which actively synthesize β -globin. These results show that the SV40 DNA fragment in plasmid 1 contains an element, the enhancer, that greatly stimulates synthesis of β -globin mRNA. [Adapted from J. Banerji et al., 1981, *Cell* **27**:299.]

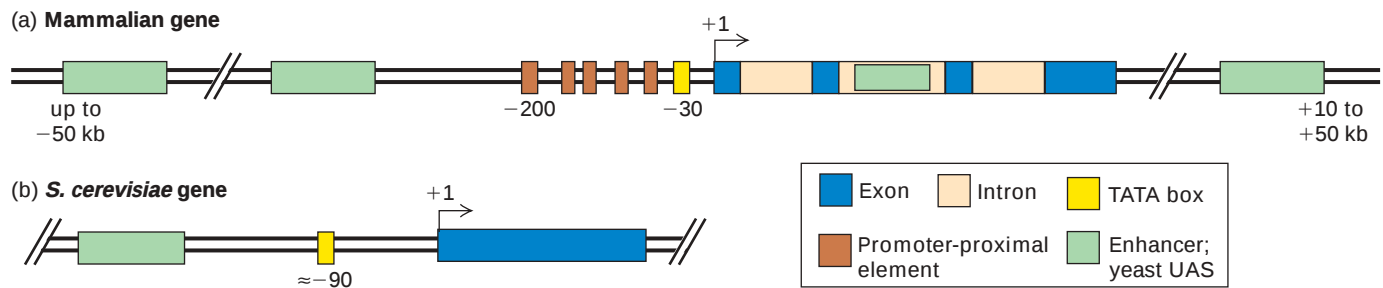
cellular DNA. Some of these control elements are located 50 or more kilobases from the promoter they control. Analyses of many different eukaryotic cellular enhancers have shown that they can occur upstream from a promoter, downstream from a promoter within an intron, or even downstream from the final exon of a gene. Like promoter-proximal elements, many enhancers are cell-type-specific. For example, the genes encoding antibodies (immunoglobulins) contain an enhancer within the second intron that can stimulate transcription from all promoters tested, but only in B lymphocytes, the type of cells that normally express antibodies. Analyses of the effects of deletions and linker scanning mutations in cellular enhancers have shown that, like the SV40 enhancer, they generally are composed of multiple elements that contribute to the overall activity of the enhancer.

Most Eukaryotic Genes Are Regulated by Multiple Transcription-Control Elements

Initially, enhancers and promoter-proximal elements were thought to be distinct types of transcription-control elements. However, as more enhancers and promoter-proximal elements were analyzed, the distinctions between them became less clear. For example, both types of element generally can stimulate transcription even when inverted, and both types often are cell-type-specific. The general consensus now is that a spectrum of control elements regulates transcription by RNA polymerase II. At one extreme are enhancers, which can stimulate transcription from a promoter tens of thousands of base pairs away (e.g., the SV40 enhancer). At the other extreme are promoter-proximal elements, such as the upstream elements controlling the HSV *tk* gene, which lose their influence when moved an additional 30–50 base pairs farther from the promoter. Researchers have identified a large number of transcription-control elements that can stimulate transcription from distances between these two extremes.

Figure 11-12a summarizes the locations of transcription-control sequences for a hypothetical mammalian gene. The start site at which transcription initiates encodes the first (5') nucleotide of the first exon of an mRNA, the nucleotide that is capped. For many genes, especially those encoding abundantly expressed proteins, a TATA box located approximately 25–35 base pairs upstream from the start site directs RNA polymerase II to begin transcription at the proper nucleotide. Promoter-proximal elements, which are relatively short (≈ 10 –20 base pairs), are located within the first ≈ 200 base pairs upstream of the start site. Enhancers, in contrast, usually are ≈ 100 base pairs long and are composed of multiple elements of ≈ 10 –20 base pairs. Enhancers may be located up to 50 kilobases upstream or downstream from the start site or within an intron. Many mammalian genes are controlled by more than one enhancer region.

The *S. cerevisiae* genome contains regulatory elements called **upstream activating sequences (UASs)**, which function



▲ FIGURE 11-12 General pattern of control elements that regulate gene expression in multicellular eukaryotes and yeast.

(a) Genes of multicellular organisms contain both promoter-proximal elements and enhancers, as well as a TATA box or other promoter element. The promoter elements position RNA polymerase II to initiate transcription at the start site and influence the rate of transcription. Enhancers may be either

upstream or downstream and as far away as 50 kb from the transcription start site. In some cases, enhancers lie within introns. For some genes, promoter-proximal elements occur downstream from the start site as well as upstream. (b) Most *S. cerevisiae* genes contain only one regulatory region, called an *upstream activating sequence* (UAS), and a TATA box, which is ≈ 90 base pairs upstream from the start site.

similarly to enhancers and promoter-proximal elements in higher eukaryotes. Most yeast genes contain only one UAS, which generally lies within a few hundred base pairs of the start site. In addition, *S. cerevisiae* genes contain a TATA box ≈ 90 base pairs upstream from the transcription start site (Figure 11-12b).

11.3 Activators and Repressors of Transcription

The various transcription-control elements found in eukaryotic DNA are binding sites for regulatory proteins. In this section, we discuss the identification, purification, and structures of these transcription factors, which function to activate or repress expression of eukaryotic protein-coding genes.

KEY CONCEPTS OF SECTION 11.2

Regulatory Sequences in Protein-Coding Genes

- Expression of eukaryotic protein-coding genes generally is regulated through multiple protein-binding control regions that are located close to or distant from the start site (Figure 11-12).
- Promoters direct binding of RNA polymerase II to DNA, determine the site of transcription initiation, and influence transcription rate.
- Three principal types of promoter sequences have been identified in eukaryotic DNA. The TATA box, the most common, is prevalent in rapidly transcribed genes. Initiator promoters are found in some genes, and CpG islands are characteristic of genes transcribed at a low rate.
- Promoter-proximal elements occur within ≈ 200 base pairs upstream of a start site. Several such elements, containing ≈ 10 – 20 base pairs, may help regulate a particular gene.
- Enhancers, which contain multiple short control elements, may be located from 200 base pairs to tens of kilobases upstream or downstream from a promoter, within an intron, or downstream from the final exon of a gene.
- Promoter-proximal elements and enhancers often are cell-type-specific, functioning only in specific differentiated cell types.

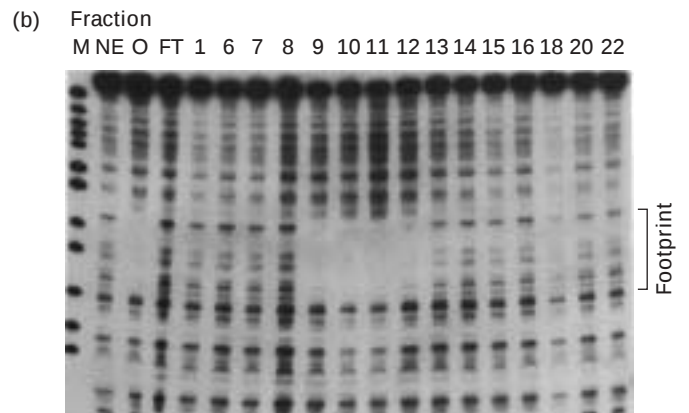
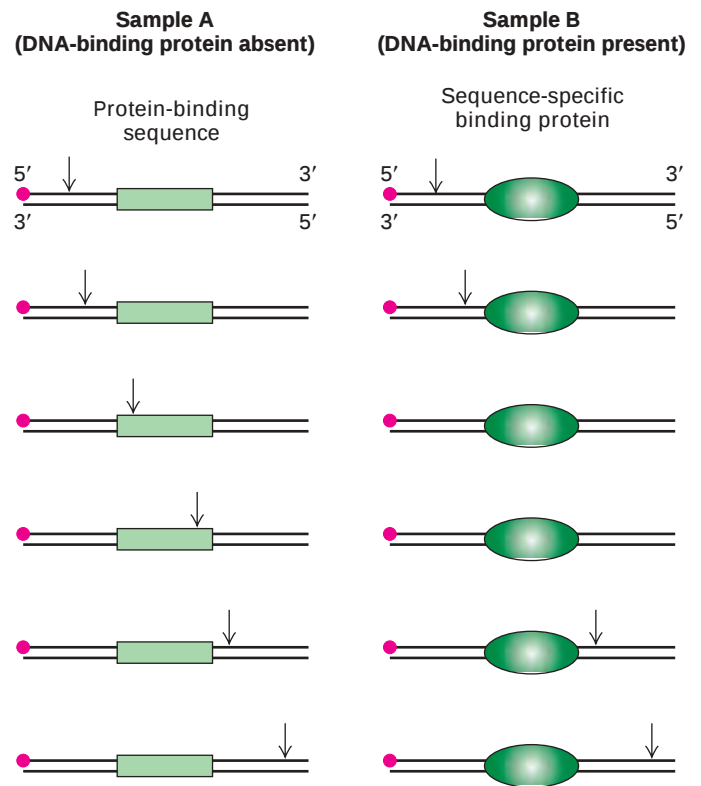
Footprinting and Gel-Shift Assays Detect Protein-DNA Interactions

In yeast, *Drosophila*, and other genetically tractable eukaryotes, numerous genes encoding transcriptional activators and repressors have been identified by classical genetic analyses like those described in Chapter 9. However, in mammals and other vertebrates, which are less amenable to such genetic analysis, most transcription factors have been detected initially and subsequently purified by biochemical techniques. In this approach, a DNA regulatory element that has been identified by the kinds of mutational analyses described in the previous section is used to identify *cognate* proteins that bind specifically to it. Two common techniques for detecting such cognate proteins are **DNase I footprinting** and the **electrophoretic mobility shift assay**.

DNase I footprinting takes advantage of the fact that when a protein is bound to a region of DNA, it protects that DNA sequence from digestion by nucleases. As illustrated in Figure 11-13, when samples of a DNA fragment that is labeled at one end are digested in the presence and absence of a DNA-binding protein and then denatured, electrophoresed, and the resulting gel subjected to autoradiography, the region protected by the bound protein appears as a gap, or “footprint,” in the array of bands resulting from digestion in the absence of protein. When footprinting is performed

► **EXPERIMENTAL FIGURE 11-13 DNase I footprinting reveals control-element sequences and can be used as an assay in transcription factor purification.**

(a) DNase I footprinting can identify control element sequences. A DNA fragment known to contain the control element is labeled at one end with ^{32}P (red dot). Portions of the labeled DNA sample then are digested with DNase I in the presence and absence of protein samples thought to contain a cognate protein. DNase I randomly hydrolyzes the phosphodiester bonds of DNA between the 3' oxygen on the deoxyribose of one nucleotide and the 5' phosphate of the next nucleotide. A low concentration of DNase I is used so that on average each DNA molecule is cleaved just once (vertical arrows). If the protein sample does not contain a cognate DNA-binding protein, the DNA fragment is cleaved at multiple positions between the labeled and unlabeled ends of the original fragment, as in sample A on the left. If the protein sample contains a cognate protein, as in sample B on the right, the protein binds to the DNA, thereby protecting a portion of the fragment from digestion. Following DNase treatment, the DNA is separated from protein, denatured to separate the strands, and electrophoresed. Autoradiography of the resulting gel detects only labeled strands and reveals fragments extending from the labeled end to the site of cleavage by DNase I. Cleavage fragments containing the control sequence show up on the gel for sample A, but are missing in sample B because the bound cognate protein blocked cleavages within that sequence and thus production of the corresponding fragments. The missing bands on the gel constitute the footprint. (b) A protein fraction containing a sequence-specific DNA-binding protein can be purified by column chromatography. DNase I footprinting can then identify which of the eluted fractions contain the cognate protein. In the absence of added protein (NE, *no extract*), DNase I cleaves the DNA fragment at multiple sites, producing multiple bands on the gel shown here. A cognate protein present in the nuclear extract applied to the column (O, *onput*) generated a footprint. This protein was bound to the column, since footprinting activity was not detected in the flow-through protein fraction (FT). After applying a salt gradient to the column, most of the cognate protein eluted in fractions 9–12, as evidenced by the missing bands (footprints). The sequence of the protein-binding region can be determined by comparison with marker DNA fragments of known length analyzed on the same gel (M). [Part (b) from S. Yoshinaga et al., 1989, *J. Biol. Chem.* **264**:10529.]



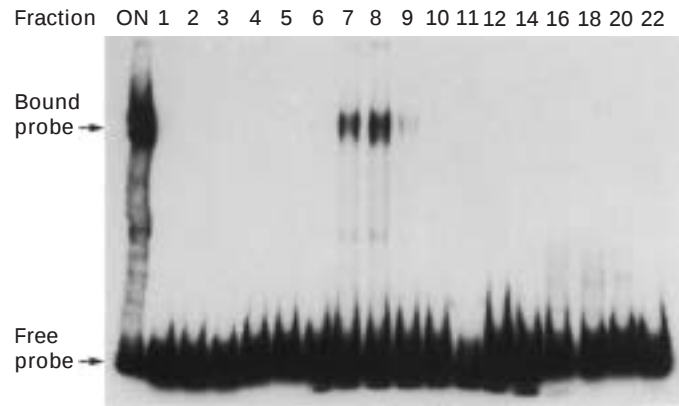
with a DNA fragment containing a known DNA control element, the appearance of a footprint indicates the presence of a transcription factor that binds that control element in the protein sample being assayed. Footprinting also identifies the specific DNA sequence to which the transcription factor binds.

The electrophoretic mobility shift assay (EMSA), also called the *gel-shift* or *band-shift* assay, is more useful than the footprinting assay for quantitative analysis of DNA-binding proteins. In general, the electrophoretic mobility of a DNA fragment is reduced when it is complexed to protein, causing a shift in the location of the fragment band. This assay can be used to detect a transcription factor in protein

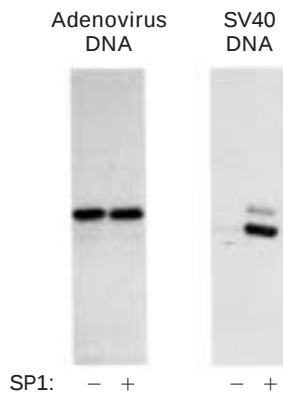
fractions incubated with a radiolabeled DNA fragment containing a known control element (Figure 11-14).

In the biochemical isolation of a transcription factor, an extract of cell nuclei commonly is subjected sequentially to several types of column chromatography (Chapter 3). Fractions eluted from the columns are assayed by DNase I footprinting or EMSA using DNA fragments containing an identified regulatory element (see Figures 11-13 and 11-14). Fractions containing protein that binds to the regulatory element in these assays probably contain a putative transcription factor. A powerful technique commonly used for the final step in purifying transcription factors is *sequence-specific DNA affinity chromatography*, a particular type of

► **EXPERIMENTAL FIGURE 11-14 Electrophoretic mobility shift assay can be used to detect transcription factors during purification.** In this example, protein fractions separated by column chromatography were assayed for their ability to bind to a radiolabeled DNA-fragment probe containing a known regulatory element. After an aliquot of the protein sample loaded onto the column (ON) and successive column fractions (numbers) were incubated with the labeled probe, the samples were electrophoresed under conditions that do not denature proteins. The free probe not bound to protein migrated to the bottom of the gel. A protein in the preparation applied to the column and in fractions 7 and 8 bound to the probe, forming a DNA-protein complex that migrated more slowly than the free probe. These fractions therefore likely contain the regulatory protein being sought. [From S. Yoshinaga et al., 1989, *J. Biol. Chem.* **264**:10529.]

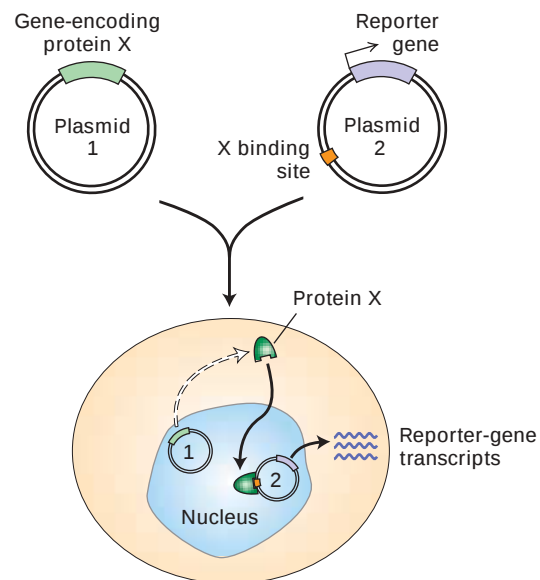


affinity chromatography in which long DNA strands containing multiple copies of the transcription factor-binding site are coupled to a column matrix. As a final test that an isolated protein is in fact a transcription factor, its ability to modulate transcription of a template containing the corresponding protein-binding sites is assayed in an *in vitro* transcription reaction. Figure 11-15 shows the results of such an assay for SP1, a transcription factor that binds to GC-rich sequences, thereby activating transcription from nearby promoters.



▲ **EXPERIMENTAL FIGURE 11-15 Transcription factors can be identified by *in vitro* assay for transcription activity.** SP1 was identified based on its ability to bind to a region of the SV40 genome that contains six copies of a GC-rich promoter-proximal element and was purified by column chromatography. To test the transcription-activating ability of purified SP1, it was incubated *in vitro* with template DNA, a protein fraction containing RNA polymerase II and associated general transcription factors, and labeled ribonucleoside triphosphates. The labeled RNA products were subjected to electrophoresis and autoradiography. Shown here are autoradiograms from assays with adenovirus and SV40 DNA in the absence (–) and presence (+) of SP1. SP1 had no significant effect on transcription from the adenovirus promoter, which contains no SP1-binding sites. In contrast, SP1 stimulated transcription from the SV40 promoter about tenfold. [Adapted from M. R. Briggs et al., 1986, *Science* **234**:47]

Once a transcription factor is isolated and purified, its partial amino acid sequence can be determined and used to clone the gene or cDNA encoding it, as outlined in Chapter 9. The isolated gene can then be used to test the ability of the encoded protein to activate or repress transcription in an *in vivo* transfection assay (Figure 11-16).



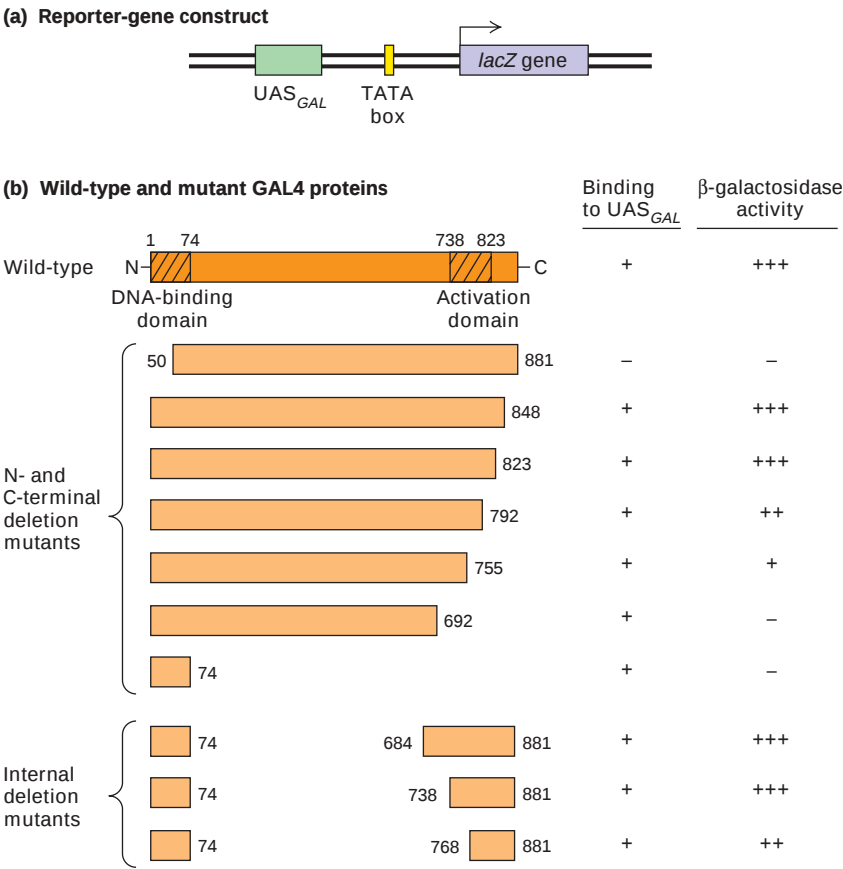
▲ **EXPERIMENTAL FIGURE 11-16 *In vivo* transfection assay measures transcription activity to evaluate proteins believed to be transcription factors.** The assay system requires two plasmids. One plasmid contains the gene encoding the putative transcription factor (protein X). The second plasmid contains a reporter gene (e.g., *lacZ*) and one or more binding sites for protein X. Both plasmids are simultaneously introduced into cells that lack the gene encoding protein X. The production of reporter-gene RNA transcripts is measured; alternatively, the activity of the encoded protein can be assayed. If reporter-gene transcription is greater in the presence of the X-encoding plasmid, then the protein is an activator; if transcription is less, then it is a repressor. By use of plasmids encoding a mutated or rearranged transcription factor, important domains of the protein can be identified.

Activators Are Modular Proteins Composed of Distinct Functional Domains

Studies with a yeast transcription activator called GAL4 provided early insight into the domain structure of transcription factors. The gene encoding the GAL4 protein, which promotes expression of enzymes needed to metabolize galactose, was identified by complementation analysis of *gal4* mutants (Chapter 9). Directed mutagenesis studies like those described previously identified UASs for the genes activated by GAL4. Each of these UASs was found to contain one or more copies of a related 17-bp sequence called UAS_{GAL}. DNase I footprinting assays with recombinant GAL4 protein produced in *E. coli* from the yeast GAL4 gene showed that

GAL4 protein binds to UAS_{GAL} sequences. When a copy of UAS_{GAL} was cloned upstream of a TATA box followed by a *lacZ* reporter gene, expression of *lacZ* was activated in galactose media in wild-type cells, but not in *gal4* mutants. These results showed that UAS_{GAL} is a transcription-control element activated by the GAL4 protein in galactose media.

A remarkable set of experiments with *gal4* deletion mutants demonstrated that the GAL4 transcription factor is composed of separable functional domains: an N-terminal **DNA-binding domain**, which binds to specific DNA sequences, and a C-terminal **activation domain**, which interacts with other proteins to stimulate transcription from a nearby promoter (Figure 11-17). When the N-terminal DNA-binding domain of GAL4 was fused directly to various



▲ EXPERIMENTAL FIGURE 11-17 Deletion mutants of the GAL4 gene in yeast with a UAS_{GAL} reporter-gene construct demonstrate the separate functional domains in an activator.

(a) Diagram of DNA construct containing a *lacZ* reporter gene and TATA box ligated to UAS_{GAL}, a regulatory element that contains several GAL4-binding sites. The reporter-gene construct and DNA encoding wild-type or mutant (deleted) GAL4 were simultaneously introduced into mutant (*gal4*) yeast cells, and the activity of β-galactosidase expressed from *lacZ* was assayed. Activity will be high if the introduced GAL4 DNA encodes a functional protein. (b) Schematic diagrams of wild-type GAL4 and various mutant forms. Small numbers refer to positions in the wild-type sequence. Deletion of 50 amino acids from the N-terminal end destroyed the

ability of GAL4 to bind to UAS_{GAL} and to stimulate expression of β-galactosidase from the reporter gene. Proteins with extensive deletions from the C-terminal end still bound to UAS_{GAL}. These results localize the DNA-binding domain to the N-terminal end of GAL4. The ability to activate β-galactosidase expression was not entirely eliminated unless somewhere between 126–189 or more amino acids were deleted from the C-terminal end. Thus the activation domain lies in the C-terminal region of GAL4. Proteins with internal deletions (*bottom*) also were able to stimulate expression of β-galactosidase, indicating that the central region of GAL4 is not crucial for its function in this assay. [See J. Ma and M. Ptashne, 1987, *Cell* 48:847; I. A. Hope and K. Struhl, 1986, *Cell* 46:885; and R. Brent and M. Ptashne, 1985, *Cell* 43:729.]

of its C-terminal fragments, the resulting truncated proteins retained the ability to stimulate expression of a reporter gene in an *in vivo* assay like that depicted in Figure 11-16. Thus the internal portion of the protein is not required for functioning of GAL4 as a transcription factor. Similar experiments with another yeast transcription factor, GCN4, which regulates genes required for synthesis of many amino acids, indicated that it contains an ≈ 60 -aa DNA-binding domain at its C-terminus and an ≈ 20 -aa activation domain near the middle of its sequence.

Further evidence for the existence of distinct activation domains in GAL4 and GCN4 came from experiments in which their activation domains were fused to a DNA-binding domain from an entirely unrelated *E. coli* DNA-binding protein. When these fusion proteins were assayed *in vivo*, they activated transcription of a reporter gene containing the cognate site for the *E. coli* protein. Thus functional transcription factors can be constructed from entirely novel combinations of prokaryotic and eukaryotic elements.

Studies such as these have now been carried out with many eukaryotic activators. The structural model of eukaryotic activators that has emerged from these studies is a modular one in which one or more activation domains are connected to a sequence-specific DNA-binding domain through flexible protein domains (Figure 11-18). In some cases, amino acids included in the DNA-binding domain also contribute to transcriptional activation. As discussed in a later section, activation domains are thought to function by

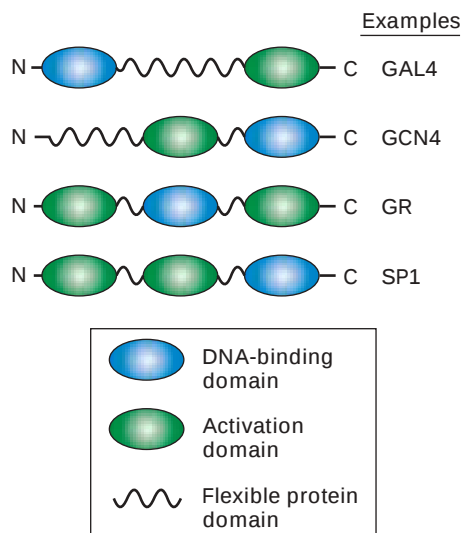
binding other proteins involved in transcription. The presence of flexible domains connecting the DNA-binding domains to activation domains may explain why alterations in the spacing between control elements are so well tolerated in eukaryotic control regions. Thus even when the positions of transcription factors bound to DNA are shifted relative to each other, their activation domains may still be able to interact because they are attached to their DNA-binding domains through flexible protein regions.

Repressors Are the Functional Converse of Activators

Eukaryotic transcription is regulated by repressors as well as activators. For example, geneticists have identified mutations in yeast that result in continuously high expression of certain genes. This type of unregulated, abnormally high expression is called **constitutive** expression and results from the inactivation of a repressor that normally inhibits the transcription of these genes. Similarly, mutants of *Drosophila* and *C. elegans* have been isolated that are defective in embryonic development because they express genes in embryonic cells where they are normally repressed. The mutations in these mutants inactivate repressors, leading to abnormal development.

Repressor-binding sites in DNA have been identified by systematic linker scanning mutation analysis similar to that depicted in Figure 11-10. In this type of analysis, mutation of an activator-binding site leads to decreased expression of the linked reporter gene, whereas mutation of a repressor-binding site leads to increased expression of a reporter gene. Repressor proteins that bind such sites can be purified and assayed using the same biochemical techniques described earlier for activator proteins.

Eukaryotic transcription repressors are the functional converse of activators. They can inhibit transcription from a gene they do not normally regulate when their cognate binding sites are placed within a few hundred base pairs of the gene's start site. Like activators, most eukaryotic repressors are modular proteins that have two functional domains: a DNA-binding domain and a **repression domain**. Similar to activation domains, repression domains continue to function when fused to another type of DNA-binding domain. If binding sites for this second DNA-binding domain are inserted within a few hundred base pairs of a promoter, expression of the fusion protein inhibits transcription from the promoter. Also like activation domains, repression domains function by interacting with other proteins as discussed later.

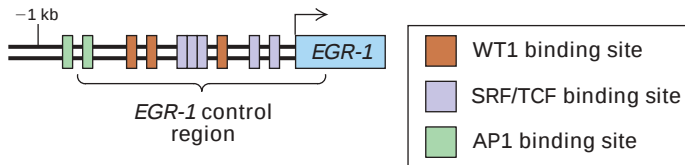


▲ FIGURE 11-18 Schematic diagrams illustrating the modular structure of eukaryotic transcription activators.

These transcription factors may contain more than one activation domain but rarely contain more than one DNA-binding domain. GAL4 and GCN4 are yeast transcription activators. The glucocorticoid receptor (GR) promotes transcription of target genes when certain hormones are bound to the C-terminal activation domain. SP1 binds to GC-rich promoter elements in a large number of mammalian genes.



The absence of appropriate repressor activity can have devastating consequences. For instance, the protein encoded by the *Wilms' tumor (WT1)* gene is a repressor that is expressed preferentially in the developing kidney. Children who inherit mutations in both the maternal and paternal *WT1* genes, so that they produce no functional WT1 protein, invariably develop kidney tumors



▲ **FIGURE 11-19 Diagram of the control region of the gene encoding *EGR-1*, a transcription activator.** The binding sites for WT1, a eukaryotic repressor protein, do not overlap the binding sites for the activator AP1 or the composite binding site for the activators SRF and TCF. Thus repression by WT1 does not involve direct interference with binding of other proteins as in the case of bacterial repressors.

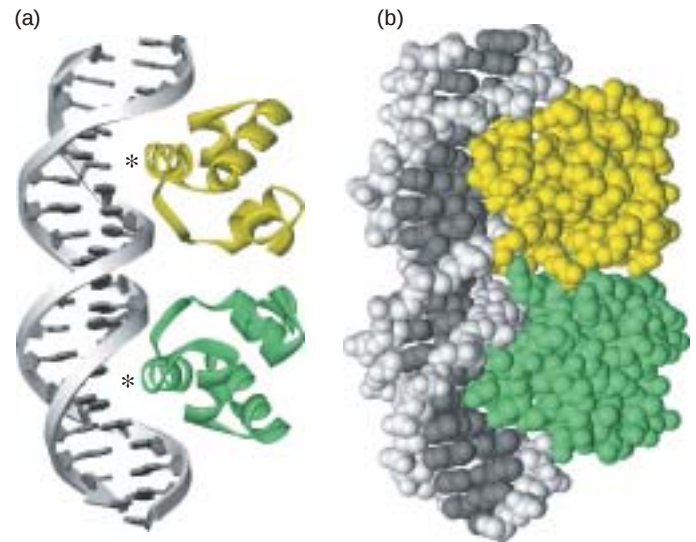
early in life. The WT1 protein binds to the control region of the gene encoding a transcription activator called *EGR-1* (Figure 11-19). This gene, like many other eukaryotic genes, is subject to both repression and activation. Binding by WT1 represses transcription of the *EGR-1* gene without inhibiting binding of the activators that normally stimulate expression of this gene. ■

DNA-Binding Domains Can Be Classified into Numerous Structural Types

The DNA-binding domains of eukaryotic activators and repressors contain a variety of structural **motifs** that bind specific DNA sequences. The ability of DNA-binding proteins to bind to specific DNA sequences commonly results from non-covalent interactions between atoms in an α helix in the DNA-binding domain and atoms on the edges of the bases within a major groove in the DNA. Interactions with sugar-phosphate backbone atoms and, in some cases, with atoms in a DNA minor groove also contribute to binding.

The principles of specific protein-DNA interactions were first discovered during the study of bacterial repressors. Many bacterial repressors are dimeric proteins in which an α helix from each monomer inserts into a major groove in the DNA helix (Figure 11-20). This α helix is referred to as the *recognition helix* or *sequence-reading helix* because most of the amino acid side chains that contact DNA extend from this helix. The recognition helix that protrudes from the surface of bacterial repressors to enter the DNA major groove and make multiple, specific interactions with atoms in the DNA is usually supported in the protein structure in part by hydrophobic interactions with a second α helix just N-terminal to it. This structural element, which is present in many bacterial repressors, is called a *helix-turn-helix* motif.

Many additional motifs that can present an α helix to the major groove of DNA are found in eukaryotic transcription factors, which often are classified according to the type of DNA-binding domain they contain. Because most of these motifs have characteristic consensus amino acid sequences, newly characterized transcription factors frequently can be classified once the corresponding genes or cDNAs are cloned



▲ **FIGURE 11-20 Interaction of bacteriophage 434 repressor with DNA.** (a) Ribbon diagram of 434 repressor bound to its specific operator DNA. Repressor monomers are in yellow and green. The recognition helices are indicated by asterisks. A space-filling model of the repressor-operator complex (b) shows how the protein interacts intimately with one side of the DNA molecule over a length of 1.5 turns. [Adapted from A. K. Aggarwal et al., 1988, *Science* **242**:899.]

and sequenced. The genomes of higher eukaryotes encode dozens of classes of DNA-binding domains and hundreds to thousands of transcription factors. The human genome, for instance, encodes ≈ 2000 transcription factors.

Here we introduce several common classes of DNA-binding proteins whose three-dimensional structures have been determined. In all these examples and many other transcription factors, at least one α helix is inserted into a major groove of DNA. However, some transcription factors contain alternative structural motifs (e.g., β strands and loops) that interact with DNA.

Homeodomain Proteins Many eukaryotic transcription factors that function during development contain a conserved 60-residue DNA-binding motif that is similar to the helix-turn-helix motif of bacterial repressors. Called **homeodomain** proteins, these transcription factors were first identified in *Drosophila* mutants in which one body part was transformed into another during development (Chapter 15). The conserved homeodomain sequence has also been found in vertebrate transcription factors, including those that have similar master control functions in human development.

Zinc-Finger Proteins A number of different eukaryotic proteins have regions that fold around a central Zn^{2+} ion, producing a compact domain from a relatively short length of the polypeptide chain. Termed a **zinc finger**, this structural motif was first recognized in DNA-binding domains but now

is known to occur also in proteins that do not bind to DNA. Here we describe two of the several classes of zinc-finger motifs that have been identified in eukaryotic transcription factors.

The C_2H_2 zinc finger is the most common DNA-binding motif encoded in the human genome and the genomes of most other multicellular animals. It is also common in multicellular plants, but is not the dominant type of DNA-binding domain in plants as it is in animals. This motif has a 23- to 26-residue consensus sequence containing two conserved cysteine (C) and two conserved histidine (H) residues, whose side chains bind one Zn^{2+} ion (see Figure 3-6b). The name “zinc finger” was coined because a two-dimensional diagram of the structure resembles a finger. When the three-dimensional structure was solved, it became clear that the binding of the Zn^{2+} ion by the two cysteine and two histidine residues folds the relatively short polypeptide sequence into a compact domain, which can insert its α helix into the major groove of DNA. Many transcription factors contain multiple C_2H_2 zinc fingers, which interact with successive groups of base pairs, within the major groove, as the protein wraps around the DNA double helix (Figure 11-21a).

A second type of zinc-finger structure, designated the C_4 zinc finger (because it has four conserved cysteines in contact with the Zn^{2+}), is found in ≈ 50 human transcription factors. The first members of this class were identified as specific intracellular high-affinity binding proteins, or “receptors,” for steroid hormones, leading to the name *steroid receptor superfamily*. Because similar intracellular receptors for non-steroid hormones subsequently were found, these transcrip-

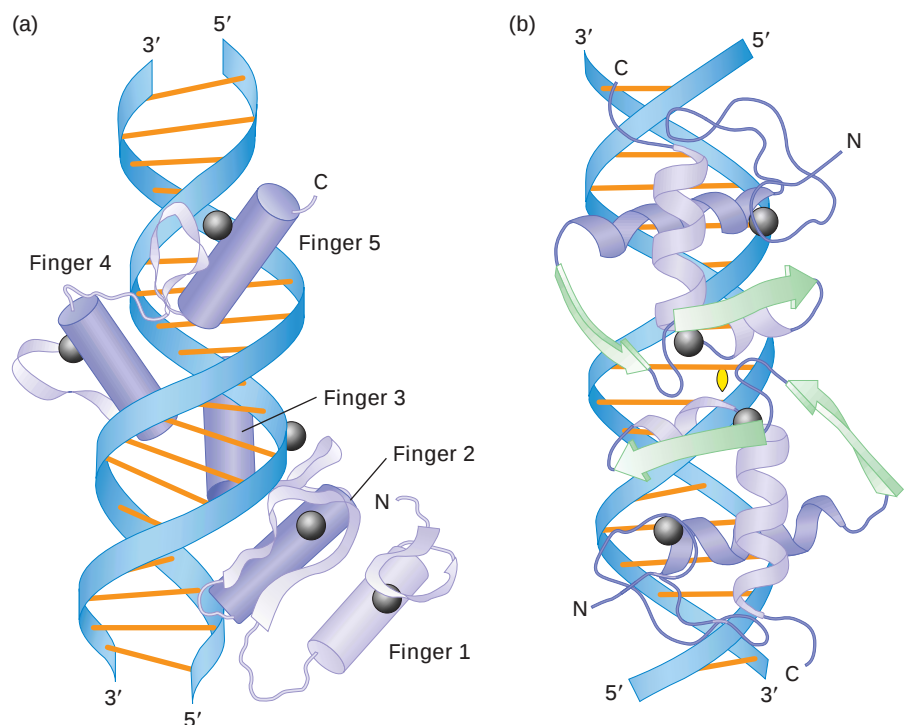
tion factors are now commonly called **nuclear receptors**. The characteristic feature of C_4 zinc fingers is the presence of two groups of four critical cysteines, one toward each end of the 55- or 56-residue domain. Although the C_4 zinc finger initially was named by analogy with the C_2H_2 zinc finger, the three-dimensional structures of proteins containing these DNA-binding motifs later were found to be quite distinct. A particularly important difference between the two is that C_2H_2 zinc-finger proteins generally contain three or more repeating finger units and bind as monomers, whereas C_4 zinc-finger proteins generally contain only two finger units and generally bind to DNA as homodimers or heterodimers. Homodimers of C_4 zinc-finger DNA-binding domains have twofold rotational symmetry (Figure 11-21b). Consequently, homodimeric nuclear receptors bind to consensus DNA sequences that are inverted repeats.

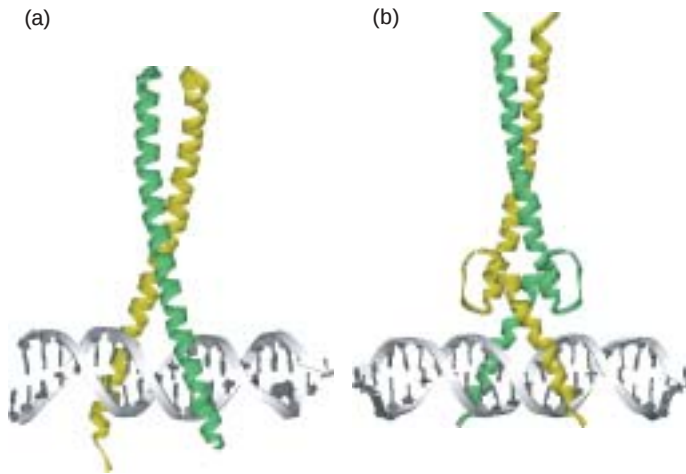
Leucine-Zipper Proteins Another structural motif present in the DNA-binding domains of a large class of transcription factors contains the hydrophobic amino acid leucine at every seventh position in the sequence. These proteins bind to DNA as dimers, and mutagenesis of the leucines showed that they were required for dimerization. Consequently, the name **leucine zipper** was coined to denote this structural motif.

The DNA-binding domain of the yeast GCN4 transcription factor mentioned earlier is a leucine-zipper domain. X-ray crystallographic analysis of complexes between DNA and the GCN4 DNA-binding domain has shown that the dimeric protein contains two extended α helices that “grip” the DNA molecule, much like a pair of scissors, at two ad-

► **FIGURE 11-21 Interaction between DNA and proteins containing zinc fingers.**

(a) GL1 is a monomeric protein that contains five C_2H_2 zinc fingers. α -Helices are shown as cylinders, Zn^{2+} ions as spheres. Finger 1 does not interact with DNA, whereas the other four fingers do. (b) The glucocorticoid receptor is a homodimeric C_4 zinc-finger protein. α -Helices are shown as purple ribbons, β -strands as green arrows, Zn^{2+} ions as spheres. Two α helices (darker shade), one in each monomer, interact with the DNA. Like all C_4 zinc-finger homodimers, this transcription factor has twofold rotational symmetry; the center of symmetry is shown by the yellow ellipse. In contrast, heterodimeric nuclear receptors do not exhibit rotational symmetry. [See N. P. Pavletich and C. O. Pabo, 1993, *Science* **261**:1701, and B. F. Luisi et al., 1991, *Nature* **352**:497.]





▲ **FIGURE 11-22 Interaction of homodimeric leucine-zipper and basic helix-loop-helix (bHLH) proteins with DNA.** (a) In leucine-zipper proteins, basic residues in the extended α -helical regions of the monomers interact with the DNA backbone at adjacent major grooves. The coiled-coil dimerization domain is stabilized by hydrophobic interactions between the monomers. (b) In bHLH proteins, the DNA-binding helices at the bottom (N-termini of the monomers) are separated by nonhelical loops from a leucine-zipper-like region containing a coiled-coil dimerization domain. [Part (a) see T. E. Ellenberger et al., 1992, *Cell* **71**:1223; part (b) see A. R. Ferre-D'Amare et al., 1993, *Nature* **363**:38.]

adjacent major grooves separated by about half a turn of the double helix (Figure 11-22a). The portions of the α helices contacting the DNA include positively charged (basic) residues that interact with phosphates in the DNA backbone and additional residues that interact with specific bases in the major groove.

GCN4 forms dimers via hydrophobic interactions between the C-terminal regions of the α helices, forming a **coiled-coil** structure. This structure is common in proteins containing amphipathic α helices in which hydrophobic amino acid residues are regularly spaced alternately three or four positions apart in the sequence, forming a stripe down one side of the α helix. These hydrophobic stripes make up the interacting surfaces between the α -helical monomers in a coiled-coil dimer (see Figure 3-6c).

Although the first leucine-zipper transcription factors to be analyzed contained leucine residues at every seventh position in the dimerization region, additional DNA-binding proteins containing other hydrophobic amino acids in these positions subsequently were identified. Like leucine-zipper proteins, they form dimers containing a C-terminal coiled-coil dimerization region and an N-terminal DNA-binding domain. The term *basic zipper* (*bZip*) now is frequently used to refer to all proteins with these common structural features. Many basic-zipper transcription factors are heterodimers of two different polypeptide chains, each containing one basic-zipper domain.

Basic Helix-Loop-Helix (bHLH) Proteins The DNA-binding domain of another class of dimeric transcription factors contains a structural motif very similar to the basic-zipper motif except that a nonhelical loop of the polypeptide chain separates two α -helical regions in each monomer (Figure 11-22b). Termed a **basic helix-loop-helix (bHLH)**, this motif was predicted from the amino acid sequences of these proteins, which contain an N-terminal α helix with basic residues that interact with DNA, a middle loop region, and a C-terminal region with hydrophobic amino acids spaced at intervals characteristic of an amphipathic α helix. As with basic-zipper proteins, different bHLH proteins can form heterodimers.

Transcription-Factor Interactions Increase Gene-Control Options

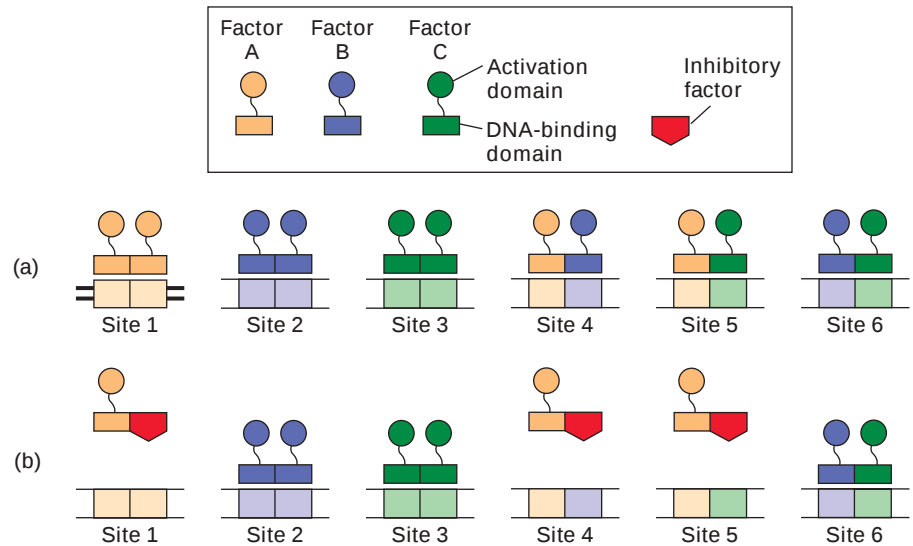
Two types of DNA-binding proteins discussed in the previous section—basic-zipper proteins and bHLH proteins—often exist in alternative heterodimeric combinations of monomers. Other classes of transcription factors not discussed here also form heterodimeric proteins. In some heterodimeric transcription factors, each monomer has a DNA-binding domain with equivalent sequence specificity. In these proteins, the formation of alternative heterodimers does not influence DNA-binding specificity, but rather allows the activation domains associated with each monomer to be brought together in alternative combinations in a single transcription factor. As we shall see later, and in subsequent chapters, the activities of individual transcription factors can be regulated by multiple mechanisms. Consequently, a single bZIP or bHLH DNA regulatory element in the control region of a gene may elicit different transcriptional responses depending on which bZIP or bHLH monomers that bind to that site are expressed in a particular cell at a particular time and how their activities are regulated.

In some heterodimeric transcription factors, however, each monomer has a different DNA-binding specificity. The resulting combinatorial possibilities increase the number of potential DNA sequences that a family of transcription factors can bind. Three different factor monomers theoretically could combine to form six homo- and heterodimeric factors, as illustrated in Figure 11-23a. Four different factor monomers could form a total of 10 dimeric factors; five monomers, 16 dimeric factors; and so forth. In addition, inhibitory factors are known that bind to some basic-zipper and bHLH monomers, thereby blocking their binding to DNA. When these inhibitory factors are expressed, they repress transcriptional activation by the factors with which they interact (Figure 11-23b). The rules governing the interactions of members of a heterodimeric transcription-factor class are complex. This combinatorial complexity expands both the number of DNA sites from which these factors can activate transcription and the ways in which they can be regulated.

Similar combinatorial transcriptional regulation is achieved through the interaction of structurally unrelated

► **FIGURE 11-23 Combinatorial possibilities due to formation of heterodimeric transcription factors.**

(a) In the hypothetical example shown, transcription factors A, B, and C can all interact with one another, permitting the three factors to bind to six different DNA sequences (sites 1–6) and creating six combinations of activation domains. Each composite binding site is divided into two half-sites, and each heterodimeric factor contains the activation domains of its two constituent monomers. (b) Expression of an inhibitory factor (green) that interacts only with factor A inhibits binding; hence, transcriptional activation at sites 1, 4, and 5 is inhibited, but activation at sites 2, 3, and 6 is unaffected.



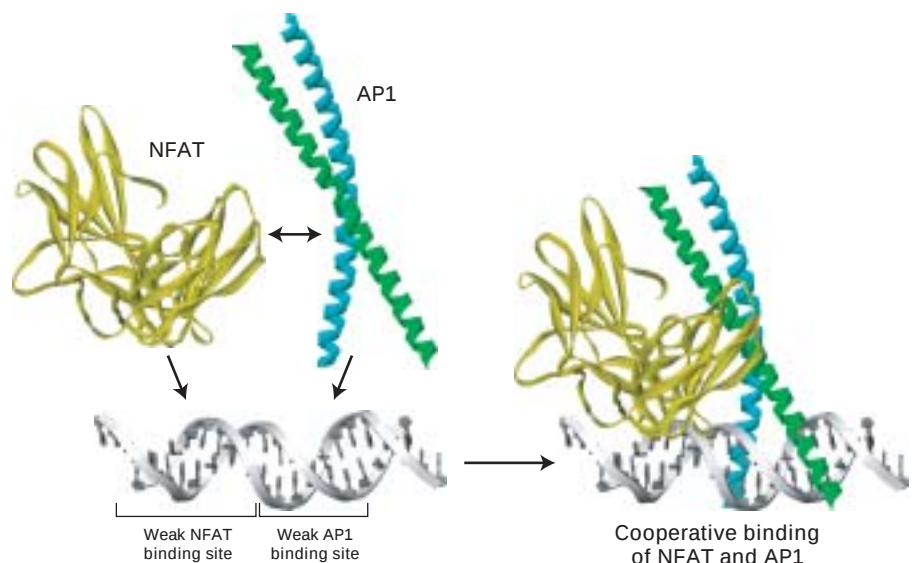
transcription factors bound to closely spaced binding sites in DNA. An example is the interaction of two transcription factors, NFAT and AP1, which bind to neighboring sites in a composite promoter-proximal element regulating the gene encoding interleukon-2 (IL-2). Expression of the *IL-2* gene is critical to the immune response, but abnormal expression of IL-2 can lead to autoimmune diseases such as rheumatoid arthritis. Neither NFAT nor AP1 binds to its site in the *IL-2* control region in the absence of the other. The affinities of the factors for these particular DNA sequences are too low for the individual factors to form a stable complex with DNA. However, when both NFAT and AP1 are present, protein-protein interactions between them stabilize the DNA ternary complex composed of NFAT, AP1, and DNA (Figure 11-24). Such *cooperative DNA binding* of various transcription factors results in considerable combinatorial complexity of transcription control. As a result, the approx-

imately 2000 transcription factors encoded in the human genome can bind to DNA through a much larger number of cooperative interactions, resulting in unique transcriptional control for each of the several tens of thousands of human genes. In the case of *IL-2*, transcription occurs only when both NFAT is activated, resulting in its transport from cytoplasm to the nucleus, and the two subunits of AP1 are synthesized. These events are controlled by distinct signal transduction pathways (Chapters 13 and 14), allowing stringent control of IL-2 expression.

Cooperative binding by NFAT and AP1 occurs only when their weak binding sites are located at a precise distance, quite close to each other in DNA. Recent studies have shown that the requirements for cooperative binding are not so stringent in the case of some other transcription factors and control regions. For example, the *EGR-1* control region contains a composite binding site to which the SRF and TCF

► **FIGURE 11-24 Cooperative binding of two unrelated transcription factors to neighboring sites in a composite control element.**

By themselves, both monomeric NFAT and heterodimeric AP1 transcription factors have low affinity for their respective binding sites in the *IL-2* promoter-proximal region. Protein-protein interactions between NFAT and AP1 add to the overall stability of the NFAT-AP1-DNA complex, so that the two proteins bind to the composite site cooperatively. [See L. Chen et al., 1998, *Nature* 392:42.]



transcription factors bind cooperatively (see Figure 11-19). Because a TCF has a long, flexible domain that interacts with SRF, the two proteins can bind cooperatively when their individual sites in DNA are separated by any distance up to 10 base pairs or are inverted relative to each other.

Structurally Diverse Activation and Repression Domains Regulate Transcription

Experiments with fusion proteins composed of the GAL4 DNA-binding domain and random segments of *E. coli* proteins demonstrated that a diverse group of amino acid sequences can function as activation domains, ~1% of all *E. coli* sequences, even though they evolved to perform other functions. Many transcription factors contain activation domains marked by an unusually high percentage of particular amino acids. GAL4, GCN4, and most other yeast transcription factors, for instance, have activation domains that are rich in acidic amino acids (aspartic and glutamic acids). These so-called *acidic activation domains* generally are capable of stimulating transcription in nearly all types of eukaryotic cells—fungal, animal, and plant cells. Activation domains from some *Drosophila* and mammalian transcription factors are glutamine-rich, and some are proline-rich; still others are rich in the closely related amino acids serine and threonine, both of which have hydroxyl groups. However, some strong activation domains are not particularly rich in any specific amino acid.

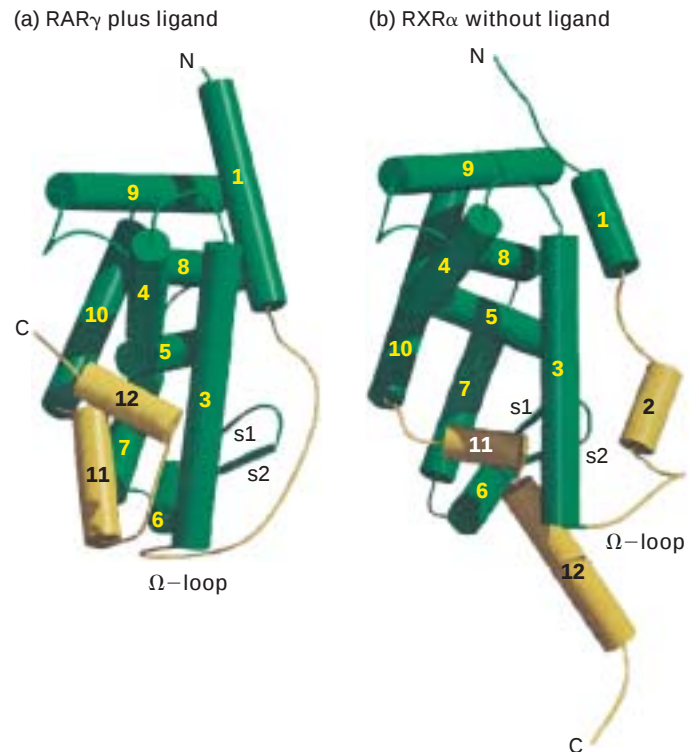
Biophysical studies indicate that acidic activation domains have an unstructured, random-coil conformation. These domains stimulate transcription when they are bound to a protein *co-activator*. The interaction with a co-activator causes the activation domain to assume a more structured α -helical conformation in the activation domain–co-activator complex. A well-studied example of a transcription factor with an acidic activation domain is the mammalian CREB protein, which is phosphorylated in response to increased levels of cAMP. This regulated phosphorylation is required for CREB to bind to its co-activator CBP (CREB binding protein), resulting in the transcription of genes whose control regions contain a CREB-binding site (see Figure 13-32). When the phosphorylated random coil activation domain of CREB interacts with CBP, it undergoes a conformational change to form two α helices that wrap around the interacting domain of CBP.

Some activation domains are larger and more highly structured than acidic activation domains. For example, the ligand-binding domains of nuclear receptors function as activation domains when they bind their specific ligand (Figure 11-25). Binding of ligand induces a large conformational change that allows the ligand-binding domain with bound hormone to interact with a short α helix in nuclear-receptor co-activators; the resulting complex then can activate transcription of genes whose control regions bind the nuclear receptor.

Thus the acidic activation domain in CREB and the ligand-binding activation domains in nuclear receptors represent

two structural extremes. The CREB acidic activation domain is a random coil that folds into two α -helices when it binds to the surface of a globular domain in a co-activator. In contrast, the nuclear-receptor ligand-binding activation domain is a structured globular domain that interacts with a short α helix in a co-activator, which probably is a random coil before it is bound. In both cases, however, specific protein-protein interactions between co-activators and the activation domains permit the transcription factors to stimulate gene expression.

Currently, less is known about the structure of repression domains. The globular ligand-binding domains of some nuclear receptors function as repression domains in the absence of their specific hormone ligand (see Figure 11-25b). Like activation domains, repression domains may be relatively short, comprising 15 or fewer amino acids. Biochemical and



▲ **FIGURE 11-25 Effect of ligand binding on conformation of the ligand-binding activation domains in two related human nuclear receptors.** Cylinders represent α helices. Regions that do not change significantly in conformation after ligand binding are shown in green; regions that do, in yellow. (a) Ligand-binding domain of human RAR γ monomer when bound to its ligand, all-*trans* retinoic acid. In this fairly compact conformation, RAR γ can stimulate transcription. (b) Ligand-binding domain of human RXR α monomer in the absence of its ligand, 9-*cis* retinoic acid. As discussed later, RXR associates with several different nuclear-receptor monomers (e.g., RAR), forming heterodimeric transcription factors. [From J. M. Wurtz et al., 1996, *Nature Struct.* **3**:87; courtesy of Hinrich Gronemeyer.]

genetic studies indicate that repression domains also mediate protein-protein interactions and bind to *co-repressor* proteins, forming a complex that inhibits transcription initiation by mechanisms that are discussed later in the chapter.

Multiprotein Complexes Form on Enhancers

As noted previously, enhancers generally range in length from about 50 to 200 base pairs and include binding sites for several transcription factors. The multiple transcription factors that bind to a single enhancer are thought to interact. Analysis of the ≈ 70 -bp enhancer that regulates expression of **β -interferon**, an important protein in defense against viral infections in humans, provides a good example of such transcription-factor interactions. The β -interferon enhancer contains four control elements that bind four different transcription factors simultaneously. In the presence of a small, abundant protein associated with chromatin called **HMGI**, binding of the transcription factors is highly cooperative, similar to the binding of NFAT and AP1 to the composite promoter-proximal site in the *IL-2* control region (see Figure 11-24). This cooperative binding produces a multiprotein complex on the β -interferon enhancer DNA (Figure 11-26). The term **enhancesome** has been coined to describe such large nucleoprotein complexes that assemble from transcription factors as they bind cooperatively to their multiple binding sites in an enhancer.

HMGI binds to the minor groove of DNA regardless of the sequence and, as a result, bends the DNA molecule sharply. This bending of the enhancer DNA permits bound transcription factors to interact properly. The inherently

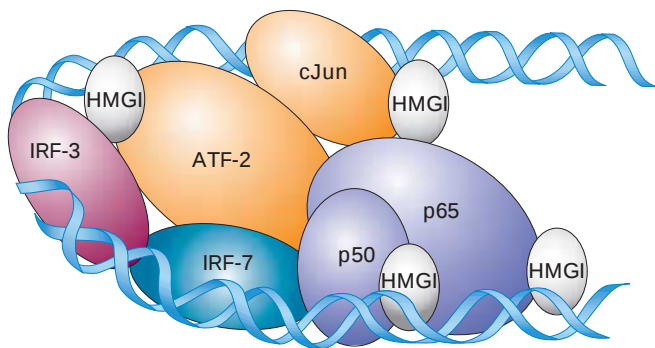
weak, noncovalent protein-protein interactions between transcription factors are strengthened by their binding to neighboring DNA sites, which keeps the proteins at very high relative concentrations.

Because of the presence of flexible regions connecting the DNA-binding domains and activation or repression domains in transcription factors (see Figure 11-18) and the ability of bound proteins to bend DNA, considerable leeway in the spacing between regulatory elements in transcription-control regions is permissible. This property probably contributed to rapid evolution of gene control in eukaryotes. Transposition of DNA sequences and recombination between repeated sequences over evolutionary time likely created new combinations of control elements that were subjected to natural selection and retained if they proved beneficial. The latitude in spacing between regulatory elements probably allowed many more functional combinations to be subjected to this evolutionary experimentation than would be the case if constraints on the spacing between regulatory elements were strict.

KEY CONCEPTS OF SECTION 11.3

Activators and Repressors of Transcription

- Transcription factors, which stimulate or repress transcription, bind to promoter-proximal regulatory elements and enhancers in eukaryotic DNA.
- Transcription activators and repressors are generally modular proteins containing a single DNA-binding domain and one or a few activation domains (for activators) or repression domains (for repressors). The different domains frequently are linked through flexible polypeptide regions (see Figure 11-18).
- Among the most common structural motifs found in the DNA-binding domains of eukaryotic transcription factors are the C_2H_2 zinc finger, homeodomain, basic helix-loop-helix (bHLH), and basic zipper (leucine zipper). All these and many other DNA-binding motifs contain one or more α helices that interact with major grooves in their cognate site in DNA.
- The transcription-control regions of most genes contain binding sites for multiple transcription factors. Transcription of such genes varies depending on the particular repertoire of transcription factors that are expressed and activated in a particular cell at a particular time.
- Combinatorial complexity in transcription control results from alternative combinations of monomers that form heterodimeric transcription factors (see Figure 11-23) and from cooperative binding of transcription factors to composite control sites (see Figure 11-24).
- Activation and repression domains in transcription factors exhibit a variety of amino acid sequences and three-dimensional structures. In general, these functional do-



▲ **FIGURE 11-26 Model of the enhancesome that forms on the β -interferon enhancer.** Two monomeric transcription factors, IRF-3 and IRF-7, and two heterodimeric factors, Jun/ATF-2 and p50/p65 (NF- κ B) bind to the four control elements in this enhancer. Cooperative binding of these transcription factors is facilitated by HMGI, which binds to the minor groove of DNA and also interacts directly with the dimeric factors. Bending of the enhancer sequence resulting from HMGI binding is critical to formation of an enhancesome. Different DNA-bending proteins act similarly at other enhancers. [Adapted from D. Thanos and T. Maniatis, 1995, *Cell* **83**:1091, and M. A. Wathel et al., 1998, *Mol. Cell* **1**:507]

mainly interact with co-activators or co-repressors, which are critical to the ability of transcription factors to modulate gene expression.

- Cooperative binding of multiple activators to nearby sites in an enhancer forms a multiprotein complex called an *enhancesome* (see Figure 11-26). Assembly of enhancesomes often requires small proteins that bind to the DNA minor groove and bend the DNA sharply, allowing bound proteins on either side of the bend to interact more readily.

11.4 Transcription Initiation by RNA Polymerase II

In previous sections many of the eukaryotic proteins and DNA sequences that participate in transcription and its control have been introduced. In this section, we focus on assembly of *transcription preinitiation complexes* involving RNA polymerase II (Pol II). Recall that this eukaryotic RNA polymerase catalyzes synthesis of mRNAs and a few small nuclear RNAs (snRNAs). Mechanisms that control the assembly of Pol II transcription preinitiation complexes, and hence the rate of transcription of protein-coding genes, are considered in the next section.

General Transcription Factors Position RNA Polymerases II at Start Sites and Assist in Initiation

In vitro transcription by purified RNA polymerase II requires the addition of several initiation factors that are separated from the polymerase during purification. These initiation factors, which position polymerase molecules at transcription start sites and help to melt the DNA strands so that the template strand can enter the active site of the enzyme, are called *general transcription factors*. In contrast to the transcription factors discussed in the previous section, which bind to specific sites in a limited number of genes, general transcription factors are required for synthesis of RNA from most genes.

The general transcription factors that assist Pol II in initiation of transcription from most TATA-box promoters in vitro have been isolated and characterized. These proteins are designated *TFIIA*, *TFIIB*, etc., and most are multimeric proteins. The largest is TFIID, which consists of a single 38-kDa *TATA box-binding protein (TBP)* and thirteen TBP-associated factors (TAFs). General transcription factors with similar activities have been isolated from cultured human cells, rat liver, *Drosophila* embryos, and yeast. The genes encoding these proteins in yeast have been sequenced as part of the complete yeast genome sequence, and many of the cDNAs encoding human and *Drosophila* Pol II general transcription factors have been cloned and sequenced. In all cases, equivalent general transcription factors from different eukaryotes are highly conserved.

Sequential Assembly of Proteins Forms the Pol II Transcription Preinitiation Complex in Vitro

Detailed biochemical studies revealed how the Pol II preinitiation complex, comprising a Pol II molecule and general transcription factors bound to a promoter region of DNA, is assembled. In these studies DNase I footprinting and electrophoretic mobility shift assays were used to determine the order in which Pol II and general transcription factors bound to TATA-box promoters. Because the complete, multisubunit TFIID is difficult to purify, researchers used only the isolated TBP component of this general transcription factor in these experiments. Pol II can initiate transcription in vitro in the absence of the other TFIID subunits.

Figure 11-27 summarizes our current understanding of the stepwise assembly of the Pol II transcription preinitiation complex in vitro. TBP is the first protein to bind to a TATA-box promoter. All eukaryotic TBPs analyzed to date have very similar C-terminal domains of 180 residues. The sequence of this region is 80 percent identical in the yeast and human proteins, and most differences are conservative substitutions. This conserved C-terminal domain functions as well as the full-length protein does in in vitro transcription. (The N-terminal domain of TBP, which varies greatly in sequence and length among different eukaryotes, functions in the Pol II-catalyzed transcription of genes encoding snRNAs.) TBP is a monomer that folds into a saddle-shape structure; the two halves of the molecule exhibit an overall dyad symmetry but are not identical. Like the HMGI and other DNA-bending proteins that participate in formation of enhancesomes, TBP interacts with the minor groove in DNA, bending the helix considerably (see Figure 4-5). The DNA-binding surface of TBP is conserved in all eukaryotes, explaining the high conservation of the TATA-box promoter element (see Figure 11-9).

Once TBP has bound to the TATA box, TFIIB can bind. TFIIB is a monomeric protein, slightly smaller than TBP. The C-terminal domain of TFIIB makes contact with both TBP and DNA on either side of the TATA-box, while its N-terminal domain extends toward the transcription start site. Following TFIIB binding, a preformed complex of tetrameric TFIIF and Pol II binds, positioning the polymerase over the start site. At most promoters, two more general transcription factors must bind before the DNA duplex can be separated to expose the template strand. First to bind is tetrameric TFIIE, creating a docking site for TFIIH, another multimeric factor containing nine subunits. Binding of TFIIH completes assembly of the transcription preinitiation complex in vitro (see Figure 11-27).

The **helicase** activity of one of the TFIIH subunits uses energy from ATP hydrolysis to unwind the DNA duplex at the start site, allowing Pol II to form an *open* complex in which the DNA duplex surrounding the start site is melted and the template strand is bound at the polymerase active site. If the remaining ribonucleoside triphosphates are present, Pol II begins transcribing the template strand. As the

polymerase transcribes away from the promoter region, another subunit of TFIIF phosphorylates the Pol II CTD at multiple sites (see Figure 11-27). In the minimal *in vitro* transcription assay containing only these general transcription factors and purified RNA polymerase II, TBP remains bound

to the TATA box as the polymerase transcribes away from the promoter region, but the other general transcription factors dissociate.

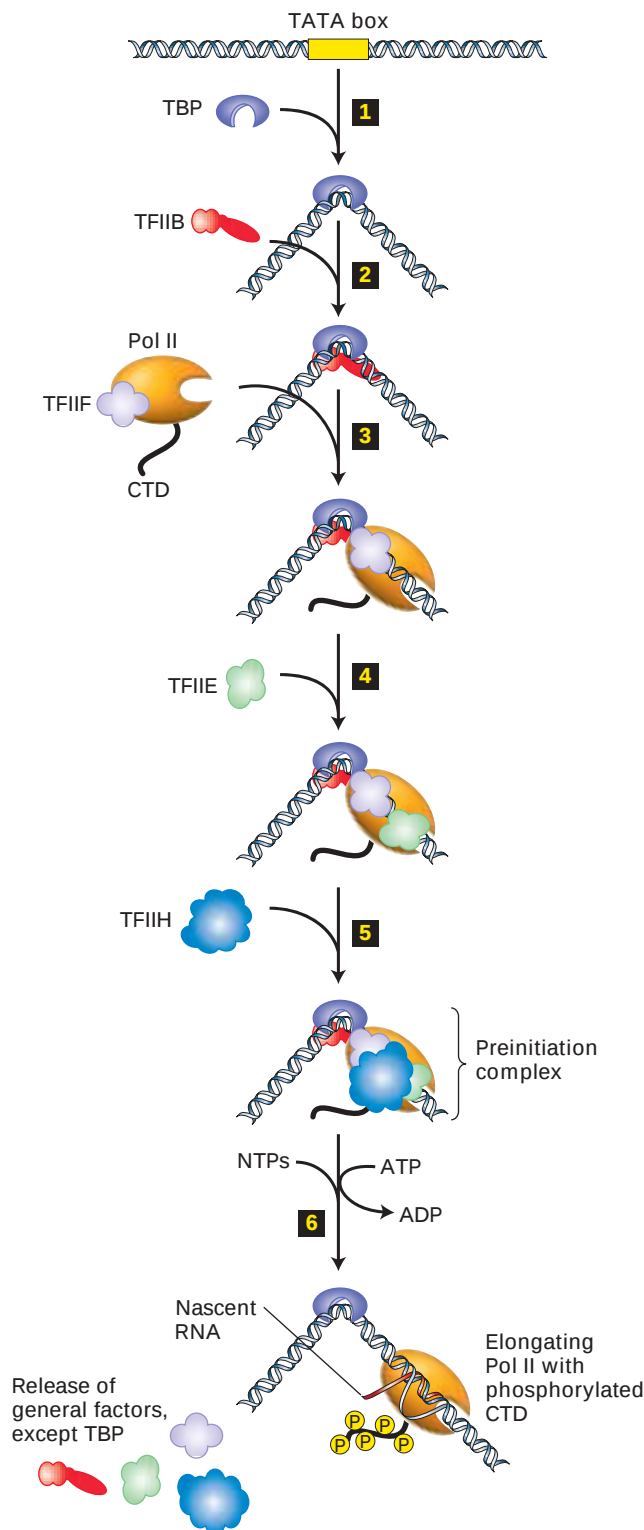


The first subunits of TFIIF to be cloned from humans were identified because mutations in the genes encoding them cause defects in the repair of damaged DNA. In normal individuals, when a transcribing RNA polymerase becomes stalled at a region of damaged template DNA, a subcomplex of TFIIF is thought to recognize the stalled polymerase and then recruit other proteins that function with TFIIF in repairing the damaged DNA region. In the absence of functional TFIIF, such repair of damaged DNA in transcriptionally active genes is impaired. As a result, affected individuals have extreme skin sensitivity to sunlight (a common cause of DNA damage) and exhibit a high incidence of cancer. Depending on the severity of the defect in TFIIF function, these individuals may suffer from diseases such as xeroderma pigmentosum and Cockayne's syndrome (Chapter 23). ■

In Vivo Transcription Initiation by Pol II Requires Additional Proteins

Although the general transcription factors discussed above allow Pol II to initiate transcription *in vitro*, another general transcription factor, TFIID, is required for initiation by Pol II *in vivo*. Purified TFIID forms a complex with TBP and TATA-box DNA. X-ray crystallography of this complex shows that TFIID interacts with the side of TBP that is upstream from the direction of transcription. Biochemical experiments suggest that in cells of higher eukaryotes TFIID and TFIIF, with its multiple TAF subunits, bind first to TATA-box DNA and then the other general transcription factors subsequently bind as indicated in Figure 11-27.

The TAF subunits of TFIID appear to play a role in initiating transcription from promoters that lack a TATA box. For instance, some TAF subunits contact the initiator element in promoters where it occurs, probably explaining how such sequences can replace a TATA box. Additional TFIID TAF subunits can bind to a consensus sequence A/G-G-A/T-C-G-T-G centered ≈30 base pairs downstream from the tran-



◀ **FIGURE 11-27 In vitro assembly of RNA polymerase II preinitiation complex.** The indicated general transcription factors and purified RNA polymerase II (Pol II) bind sequentially to TATA-box DNA to form a preinitiation complex. ATP hydrolysis then provides the energy for unwinding of DNA at the start site by a TFIIF subunit. As Pol II initiates transcription in the resulting open complex, the polymerase moves away from the promoter and its CTD becomes phosphorylated. *In vitro*, the general transcription factors (except for TBP) dissociate from the TBP-promoter complex, but it is not yet known which factors remain associated with promoter regions following each round of transcription initiation *in vivo*.

scription start site in many genes that lack a TATA-box promoter. Because of its position, this regulatory sequence is called the *downstream promoter element (DPE)*. The DPE facilitates transcription of TATA-less genes that contain it by increasing TFIID binding.

In addition to general transcription factors, specific activators and repressors regulate transcription of genes by Pol II. In the next section we will examine how these regulatory proteins influence Pol II transcription initiation.

KEY CONCEPTS OF SECTION 11.4

Transcription Initiation by RNA Polymerase II

- Transcription of protein-coding genes by Pol II can be initiated *in vitro* by sequential binding of the following in the indicated order: TBP, which binds to TATA-box DNA; TFIIB; a complex of Pol II and TFIIF; TFIIE; and finally TFIIH (see Figure 11-27).
- The helicase activity of a TFIIH subunit separates the template strands at the start site in most promoters, a process that requires hydrolysis of ATP. As Pol II begins transcribing away from the start site, its CTD is phosphorylated by another TFIIH subunit.
- *In vivo* transcription initiation by Pol II also requires TFIIA and, in metazoans, a complete TFIID protein, including its multiple TAF subunits as well as the TBP subunit.

11.5 Molecular Mechanisms of Transcription Activation and Repression

The activators and repressors that bind to specific sites in DNA and regulate expression of the associated protein-coding genes do so by two general mechanisms. First, these regulatory proteins act in concert with other proteins to modulate chromatin structure, thereby influencing the ability of general transcription factors to bind to promoters. Recall from Chapter 10 that the DNA in eukaryotic cells is not free, but is associated with a roughly equal mass of protein in the form of **chromatin**. The basic structural unit of chromatin is the **nucleosome**, which is composed of ≈ 147 base pairs of DNA wrapped tightly around a disk-shaped core of **histone** proteins. Residues within the N-terminal region of each histone, and the C-terminal region of histone H2A, called *histone tails*, extend from the surface of the nucleosome, and can be reversibly modified (see Figure 10-20). Such modifications, especially the acetylation of histone H3 and H4 tails, influence the relative condensation of chromatin and thus its accessibility to proteins required for transcription initiation. In addition to their role in such chromatin-mediated transcriptional control, activators and repressors interact with a large multiprotein complex called the *mediator of transcription complex*, or simply **mediator**.

This complex in turn binds to Pol II and directly regulates assembly of transcription preinitiation complexes.

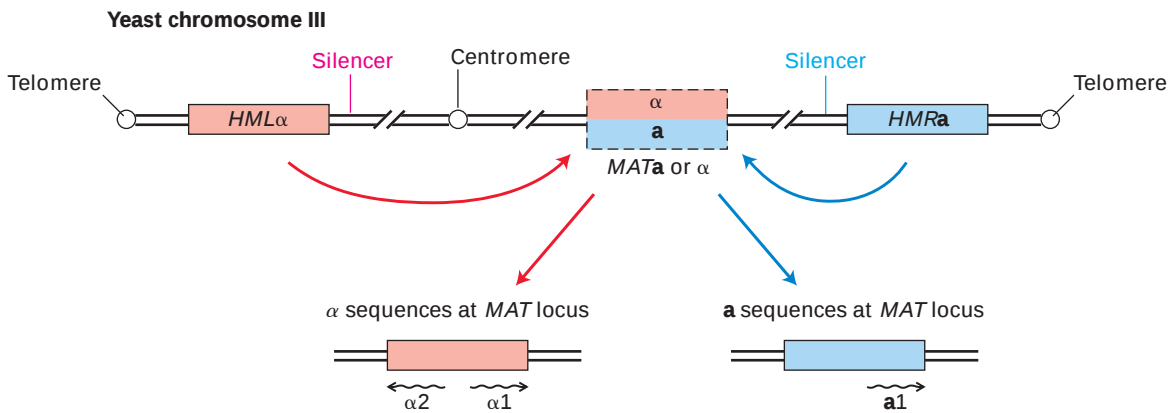
In this section, we review current understanding of how activators and repressors control chromatin structure and preinitiation complex assembly. In the next section of the chapter, we discuss how the concentrations and activities of activators and repressors themselves are controlled, so that gene expression is precisely attuned to the needs of the cell and organism.

Formation of Heterochromatin Silences Gene Expression at Telomeres, near Centromeres, and in Other Regions

For many years it has been clear that inactive genes in eukaryotic cells are often associated with **heterochromatin**, regions of chromatin that are more highly condensed and stain more darkly with DNA dyes than **euchromatin**, where most transcribed genes are located (see Figure 5-25). Regions of chromosomes near the centromeres and telomeres and additional specific regions that vary in different cell types are organized into heterochromatin. The DNA in heterochromatin is less accessible to externally added proteins than DNA in euchromatin, and consequently often referred to as “closed” chromatin. For instance, in an experiment described in the last chapter, the DNA of inactive genes was found to be far more resistant to digestion by DNase I than the DNA of transcribed genes (see Figure 10-22).

Study of DNA regions in *S. cerevisiae* that behave like the heterochromatin of higher eukaryotes provided early insight into the *chromatin-mediated repression* of transcription. This yeast can grow either as haploid or diploid cells. Haploid cells exhibit one of two possible mating types, called **a** and α . Cells of different mating type can “mate,” or fuse, to generate a diploid cell (see Figure 1-5). When a haploid cell divides by budding, the larger “mother” cell switches its mating type (see Figure 22-21). Genetic and molecular analyses have revealed that three genetic loci on yeast chromosome III control the mating type of yeast cells (Figure 11-28). Only the central mating-type locus, termed *MAT*, is actively transcribed. How the proteins encoded at the *MAT* locus determine whether a cell has the **a** or α phenotype is explained in Chapter 22. The two additional loci, termed *HML* and *HMR*, near the left and right telomere, respectively, contain “silent” (nontranscribed) copies of the **a** or α genes. These sequences are transferred alternately from *HML* α or *HMR***a** into the *MAT* locus by a type of nonreciprocal recombination between sister chromatids during cell division. When the *MAT* locus contains the DNA sequence from *HML* α , the cells behave as α cells. When the *MAT* locus contains the DNA sequence from *HMR***a**, the cells behave like **a** cells.

Our interest here is how transcription of the silent mating-type loci at *HML* and *HMR* are repressed. If the genes at these loci are expressed, as they are in yeast mutants with defects in the repressing mechanism, both **a** and α proteins are expressed, causing the cells to behave like diploid cells,



▲ **FIGURE 11-28 Arrangement of mating-type loci on chromosome III in the yeast *S. cerevisiae*.** Silent (unexpressed) mating-type genes (either *a* or α , depending on the strain) are located at the *HML* locus. The opposite mating-type genes are present at the silent *HMR* locus. When the α or *a* sequences are present at the *MAT* locus, they can be transcribed into mRNAs

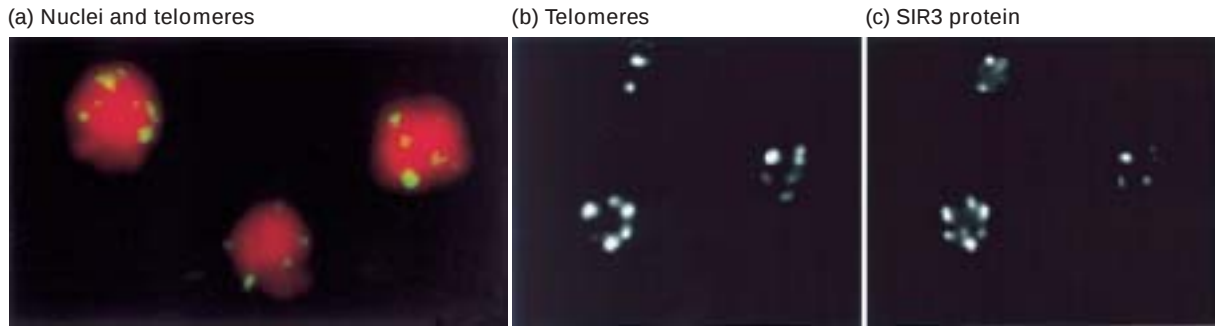
whose encoded proteins specify the mating-type phenotype of the cell. The silencer sequences near *HML* and *HMR* bind proteins that are critical for repression of these silent loci. Haploid cells can switch mating types in a process that transfers the DNA sequence from *HML* or *HMR* to the transcriptionally active *MAT* locus.

which cannot mate. The promoters and UASs controlling transcription of the *a* and α genes lie near the center of the DNA sequence that is transferred and are identical whether the sequences are at the *MAT* locus or at one of the silent loci. Consequently, the function of the transcription factors that interact with these sequences is somehow blocked at *HML* and *HMR*. This repression of the silent loci depends on **silencer sequences** located next to the region of transferred DNA at *HML* and *HMR* (see Figure 11-28). If the silencer is deleted, the adjacent silent locus is transcribed. Remarkably, any gene placed near the yeast mating-type silencer sequence by recombinant DNA techniques is repressed, or “silenced,” even a tRNA gene transcribed by RNA polymerase III, which uses a different set of general transcription factors than RNA polymerase II uses.

Several lines of evidence indicate that repression of the *HML* and *HMR* loci results from a condensed chromatin structure that sterically blocks transcription factors from interacting with the DNA. In one telling experiment, the gene encoding an *E. coli* enzyme that methylates adenine residues in GATC sequences was introduced into yeast cells under the control of a yeast promoter so that the enzyme was expressed. Researchers found that GATC sequences within the *MAT* locus and most other regions of the genome in these cells were methylated, but not those within the *HML* and *HMR* loci. These results indicate that the DNA of the silent loci is inaccessible to the *E. coli* methylase and presumably to proteins in general, including transcription factors and RNA polymerase. Similar experiments conducted with various yeast histone mutants indicated that specific interactions involving the histone tails of H3 and H4 are required for formation of a fully repressing chromatin structure. Other studies have shown that the telomeres of every yeast chromosome also behave like silencer sequences. For instance,

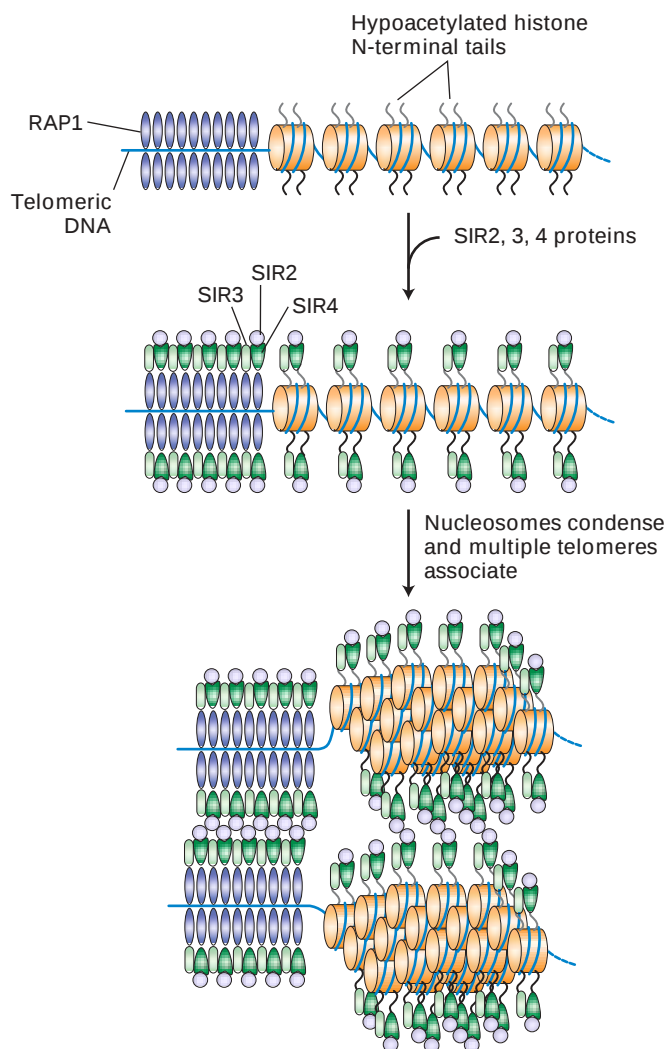
when a gene is placed within a few kilobases of any yeast telomere, its expression is repressed. In addition, this repression is relieved by the same mutations in the H3 and H4 histone tails that interfere with repression at the silent mating-type loci.

Genetic studies led to identification of several proteins, RAP1 and three SIR proteins, that are required for repression of the silent mating-type loci and the telomeres in yeast. RAP1 was found to bind within the DNA silencer sequences associated with *HML* and *HMR* and to a sequence that is repeated multiple times at each yeast chromosome telomere. Further biochemical studies showed that the SIR proteins bind to one another and that two bind to the N-terminal tails of histones H3 and H4 that are maintained in a largely unacetylated state by the deacetylase activity of SIR2. Several experiments using fluorescence confocal microscopy of yeast cells either stained with fluorescent-labeled antibody to any one of the SIR proteins or RAP1 or hybridized to a labeled telomere-specific DNA probe revealed that these proteins form large, condensed telomeric nucleoprotein structures resembling the heterochromatin found in higher eukaryotes (Figure 11-29). Figure 11-30 depicts a model for the chromatin-mediated silencing at yeast telomeres based on these and other studies. Formation of heterochromatin at telomeres is nucleated by multiple RAP1 proteins bound to repeated sequences in a nucleosome-free region at the extreme end of a telomere. A network of protein-protein interactions involving telomere-bound RAP1, three SIR proteins (2, 3, and 4), and hypo-acetylated histones H3 and H4 creates a stable, higher-order nucleoprotein complex that includes several telomeres and in which the DNA is largely inaccessible to external proteins. One additional protein, SIR1, is also required for silencing of the silent mating-type loci. Although the function of SIR1 is not yet well understood, it is known



▲ **EXPERIMENTAL FIGURE 11-29 Antibody and DNA probes colocalize SIR3 protein with telomeric heterochromatin in yeast nuclei.** (a) Confocal micrograph 0.3 μm thick through three diploid yeast cells, each containing 68 telomeres. Telomeres were labeled by hybridization to a fluorescent telomere-specific probe (yellow). DNA was stained red to reveal the nuclei. The 68 telomeres coalesce into a much smaller number of regions near the nuclear periphery. (b, c)

Confocal micrographs of yeast cells labeled with a telomere-specific hybridization probe (b) and a fluorescent-labeled antibody specific for SIR3 (c). Note that SIR3 is localized in the repressed telomeric heterochromatin. Similar experiments with RAP1, SIR2, and SIR4 have shown that these proteins also colocalize with the repressed telomeric heterochromatin. [From M. Gotta et al., 1996, *J. Cell Biol.* **134**:1349; courtesy of M. Gotta, T. Laroche, and S. M. Gasser.]



to associate with the silencer region, where it is thought to cause further assembly of the multiprotein telomeric silencing complex so that it spreads farther from the end of the chromosome, encompassing *HML* and *HMR*.

An important feature of this model is the dependence of silencing on hypoacetylation of the histone tails. This was shown in experiments with yeast mutants expressing histones in which lysines in histone N-termini were substituted with arginines or glutamines. Arginine is positively charged like lysine, but cannot be acetylated. It is thought to function in histone N-terminal tails like an unacetylated lysine. Glutamine on the other hand simulates an acetylated lysine. Repression at telomeres and at the silent mating-type loci was defective in the mutants with glutamine substitutions, but not in mutants with arginine substitutions. Hyperacetylation of the H3 and H4 tails subsequently was found to interfere with binding by SIR3 and SIR4.

Although chromatin-mediated repression of transcription is also important in multicellular eukaryotes, the mechanism of this repression is still being worked out. Genetic and

◀ **FIGURE 11-30 Schematic model of silencing mechanism at yeast telomeres.** Multiple copies of RAP1 bind to a simple repeated sequence at each telomere region, which lacks nucleosomes (*top*). This nucleates the assembly of a multiprotein complex (*bottom*) through protein-protein interactions between RAP1, SIR2, SIR3, SIR4, and the hypoacetylated N-terminal tails of histones H3 and H4 of nearby nucleosomes. SIR2 deacetylates the histone tails. The heterochromatin structure at each telomere encompasses ≈ 4 kb of DNA neighboring the RAP1-binding sites, irrespective of its sequence. Association of several condensed telomeres forms higher-order heterochromatin complexes, such as those shown in Figure 11-29, that sterically block other proteins from interacting with the DNA. See the text for more details. [Adapted from M. Grunstein, 1997, *Curr. Opin. Cell Biol.* **9**:383.]

biochemical analyses in *Drosophila* have revealed that multiple proteins associated in large multiprotein complexes participate in the process. These *Polycomb complexes* are discussed further in Chapter 15. As in the case of SIR proteins binding to yeast telomeres (see Figure 11-29), these *Drosophila* Polycomb proteins can be visualized binding to the genes they repress at multiple, specific locations in the genome by in situ binding of specific-labeled antibodies to salivary gland polytene chromosomes.

Repressors Can Direct Histone Deacetylation at Specific Genes

The importance of *histone deacetylation* in chromatin-mediated gene repression has been further supported by studies of eukaryotic repressors that regulate genes at internal chromosomal positions. These proteins are now known to act in part by causing deacetylation of histone tails in nucleosomes that bind to the TATA box and promoter-proximal region of the genes they repress. In vitro studies have shown that when promoter DNA is assembled onto a nucleosome with unacetylated histones, the general transcription factors cannot bind to the TATA box and initiation region. In un-

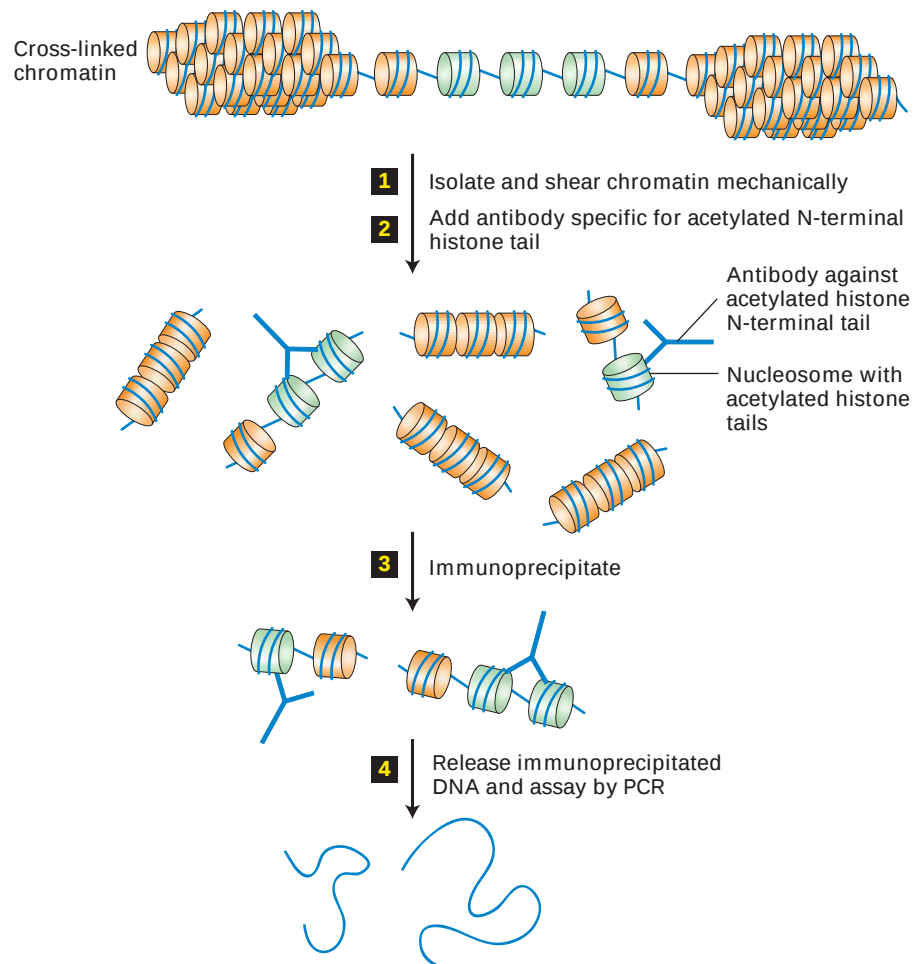
acetylated histones, the N-terminal lysines are positively charged and interact strongly with DNA phosphates. The unacetylated histone tails also interact with neighboring histone octamers, favoring the folding of chromatin into condensed, higher-order structures whose precise conformation is not well understood. The net effect is that general transcription factors cannot assemble into a preinitiation complex on a promoter associated with hypoacetylated histones. In contrast, binding of general transcription factors is repressed much less by histones with hyperacetylated tails in which the positively charged lysines are neutralized and electrostatic interactions with DNA phosphates are eliminated.

The connection between histone deacetylation and repression of transcription at nearby yeast promoters became clearer when the cDNA encoding a human *histone deacetylase* was found to have high homology to the yeast *RPD3* gene, known to be required for the normal repression of a number of yeast genes. Further work showed that RPD3 protein has histone deacetylase activity. The ability of RPD3 to deacetylate histones at a number of promoters depends on two other proteins: UME6, a repressor that binds to a specific upstream regulatory sequence (URS1), and SIN3, which is part of a large, multiprotein complex that also contains

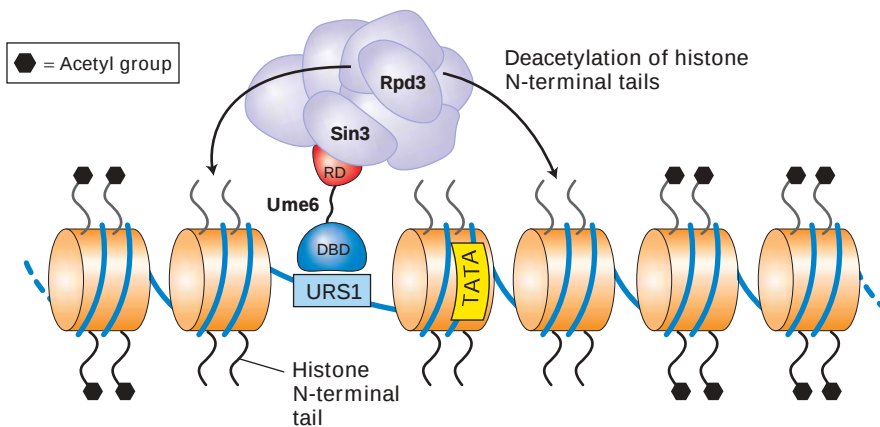
► EXPERIMENTAL FIGURE 11-31

The chromatin immunoprecipitation method can reveal the acetylation state of histones in chromatin.

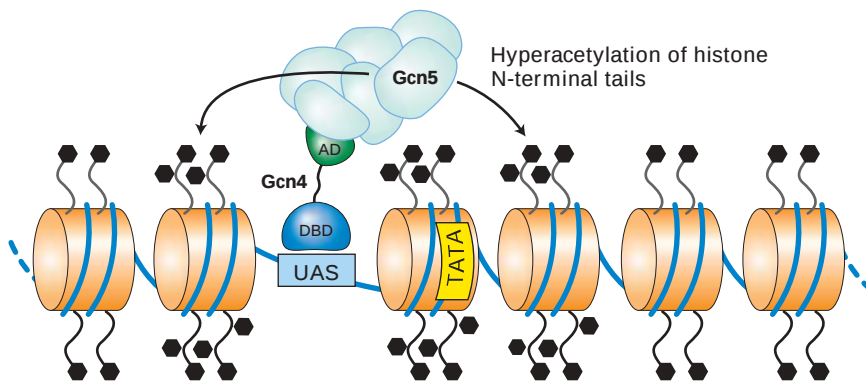
Histones are lightly cross-linked to DNA in vivo using a cell-permeable, reversible, chemical cross-linking agent. Nucleosomes with acetylated histone tails are shown in green. Step **1**: Cross-linked chromatin is then isolated and sheared to an average length of two to three nucleosomes. Step **2**: An antibody against a particular acetylated histone tail sequence is added, and (step **3**) bound nucleosomes are immunoprecipitated. Step **4**: DNA in the immunoprecipitated chromatin fragments is released by reversing the cross-link and then is quantitated using a sensitive PCR method. The method can be used to analyze the in vivo association of any protein with a specific sequence of DNA by using an antibody against the protein of interest in step **2**. [See S. E. Rundlett et al., 1998, *Nature* **392**:831]



(a) Repressor-directed histone deacetylation



(b) Activator-directed histone hyperacetylation



◀ **FIGURE 11-32 Proposed mechanism of histone deacetylation and hyperacetylation in yeast transcription control.** (a) Repressor-directed deacetylation of histone N-terminal tails. The DNA-binding domain (DBD) of the repressor UME6 interacts with a specific upstream control element (URS1) of the genes it regulates. The UME6 repression domain (RD) binds SIN3, a subunit of a multiprotein complex that includes RPD3, a histone deacetylase. Deacetylation of histone N-terminal tails on nucleosomes in the region of the UME6-binding site inhibits binding of general transcription factors at the TATA box, thereby repressing gene expression. (b) Activator-directed hyperacetylation of histone N-terminal tails. The DNA-binding domain of the activator GCN4 interacts with specific upstream activating sequences (UAS) of the genes it regulates. The GCN4 activation domain (AD) then interacts with a multiprotein histone acetylase complex that includes the GCN5 catalytic subunit. Subsequent hyperacetylation of histone N-terminal tails on nucleosomes in the vicinity of the GCN4-binding site facilitates access of the general transcription factors required for initiation. Repression and activation of many genes in higher eukaryotes occurs by similar mechanisms.

RPD3. SIN3 also binds to the repression domain of UME6, thus positioning the RPD3 histone deacetylase in the complex so it can interact with nearby promoter-associated nucleosomes and remove acetyl groups from histone N-terminal lysines. Additional experiments, using the *chromatin immunoprecipitation* technique outlined in Figure 11-31, demonstrated that in wild-type yeast, one or two nucleosomes in the immediate vicinity of UME6-binding sites are hypoacetylated. These DNA regions include the promoters of genes repressed by UME6. In *sin3* and *rp3* deletion mutants, not only were these promoters derepressed, but the nucleosomes near the UME6-binding sites were hyperacetylated.

All these findings provide considerable support for the model of repressor-directed deacetylation shown in Figure 11-32a. In this model, the SIN3-RPD3 complex functions as a co-repressor. Co-repressor complexes containing histone deacetylases also have been found associated with many repressors from mammalian cells. Some of these complexes contain the mammalian homolog of SIN3 (mSin3), which interacts with the repressor protein. Other histone deacetylase complexes identified in mammalian cells appear to contain additional or different repressor-binding proteins. These various repressor and co-repressor combinations are thought to mediate histone deacetylation at specific promoters by a mechanism similar to the yeast mechanism (see

Figure 11-32a). However, the observation that a number of eukaryotic repressor proteins inhibit *in vitro* transcription in the absence of histones indicates that more direct repression mechanisms, not involving histone deacetylation, also operate.

The discovery of mSin3-containing histone deacetylase complexes provides an explanation for earlier observations that in vertebrates transcriptionally inactive DNA regions often contain the modified cytidine residue 5-methylcytidine (mC) followed immediately by a G, whereas transcriptionally active DNA regions lack mC residues. DNA containing 5-methylcytidine has been found to bind a specific protein that in turn interacts specifically with mSin3. This finding suggests that association of mSin3-containing co-repressors with methylated sites in DNA leads to deacetylation of histones in neighboring nucleosomes, making these regions inaccessible to general transcription factors and Pol II, and hence transcriptionally inactive.

Activators Can Direct Histone Acetylation at Specific Genes

Genetic and biochemical studies in yeast led to discovery of a large multiprotein complex containing the protein GCN5, which has histone acetylase activity. Another subunit of this

histone acetylase complex binds to acidic activation domains in yeast activator proteins such as GCN4. Maximal transcription activation by GCN4 depends on these histone acetylase complexes, which thus function as co-activators. The model shown in Figure 11-32b is consistent with the observation that nucleosomes near the promoter region of a gene regulated by the GCN4 activator are specifically hyperacetylated, as determined by the chromatin immunoprecipitation method. The activator-directed hyperacetylation of nucleosomes near a promoter region changes (“opens”) the chromatin structure so as to facilitate the binding of other proteins required for transcription initiation.

A similar activation mechanism operates in higher eukaryotes. For example, mammals express two related ≈400-kD, multidomain proteins called *CBP* and *P300*, which are thought to function similarly. As noted earlier, one domain of CBP binds the phosphorylated acidic activation domain in the CREB transcription factor. Other domains of CBP interact with different activation domains in other transcription factors. Yet another domain of CBP has histone acetylase activity, and another CBP domain associates with a multiprotein histone acetylase complex that is homologous to the yeast GCN5-containing complex. CREB and many other mammalian activators are thought to function in part by directing CBP and the associated histone acetylase complex to specific nucleosomes, where they acetylate histone tails, facilitating the interaction of general transcription factors with promoter DNA. In addition, the largest TFIID subunit also has histone acetylase activity and may function as a co-activator by acetylating histone N-terminal tails in the vicinity of the TATA box.

Modifications of Specific Residues in Histone Tails Control Chromatin Condensation

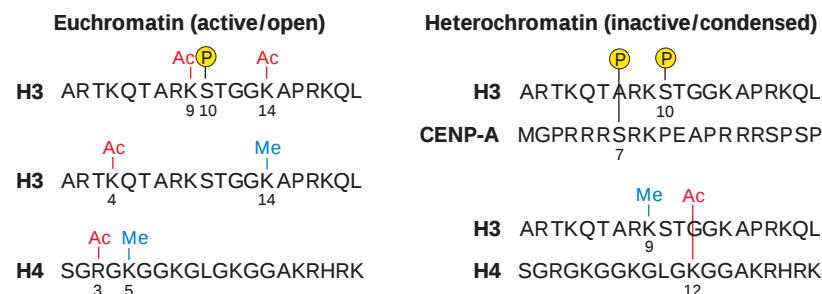
In addition to reversible acetylation, histone tails in chromatin can undergo reversible phosphorylation of serine and threonine residues, reversible monoubiquitination of a lysine residue in the H2A C-terminal tail, and irreversible methy-

lation of lysine residues. Evidence is accumulating that it is not simply the overall level of histone acetylation that controls the condensation of chromatin and hence the accessibility of DNA. Rather the precise amino acids in the tails that are acetylated or otherwise modified may constitute a “*histone code*” that helps control the condensation of chromatin (Figure 11-33). For instance, the lysine at position 9 in histone H3 often is methylated in heterochromatin.

The histone code is “read” by proteins that bind to these specific modifications and in turn promote condensation or decondensation of chromatin, forming “closed” or “open” chromatin structures. For example, higher eukaryotes express a number of heterochromatin-associated proteins containing a so-called *chromodomain*, which binds to the histone H3 tail when it is methylated at lysine 9. These proteins are postulated to contribute to the higher-order folding characteristic of heterochromatin, somewhat like the SIR proteins at yeast telomeres (see Figure 11-30). Alternatively, the *bromodomain* found in a number of euchromatin-associated proteins binds to acetylated histone tails. The largest subunit of TFIID, for example, contains two closely spaced bromodomains, which may help it to associate with chromatin containing an active code, while the histone acetylase activity of this same subunit maintains the chromatin in a hyperacetylated state.

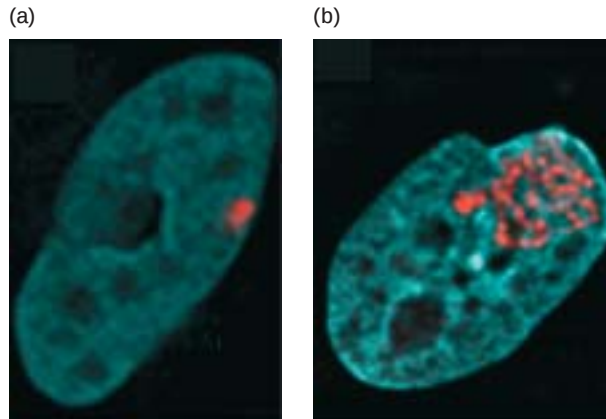
Chromatin-Remodeling Factors Help Activate or Repress Some Genes

In addition to histone acetylase complexes, another type of multiprotein complex, called the *SWI/SNF chromatin-remodeling complex*, is required for activation at some yeast promoters. Several of the SWI/SNF subunits have homology to DNA helicases, enzymes that use energy from ATP hydrolysis to disrupt interactions between base-paired nucleic acids or between nucleic acids and proteins. The SWI/SNF complex is thought to transiently dissociate DNA from the surface of nucleosomes, permitting nucleosomes to “slide” along the DNA and promoting the unfolding of condensed,



▲ **FIGURE 11-33 Examples of the histone code.** Specific post-translational modifications of the N-terminal tails in histones H3 and H4 are found in euchromatin, which is accessible to proteins and transcriptionally active. Different modifications are found in heterochromatin, which is condensed and thus largely

inaccessible to proteins and transcriptionally inactive. Histone tail sequences are shown in the one-letter amino acid code. CENP-A is a variant form for H3 found in nucleosomes associated with the centromeres of mammalian chromosomes. [Adapted from T. Jenuwein and C. D. Allis, 2001, *Science* **293**:1074.]



▲ EXPERIMENTAL FIGURE 11-34 Expression of fusion proteins demonstrates chromatin decondensation in response to an activation domain. A cultured hamster cell line was engineered to contain multiple copies of a tandem array of *E. coli lac operator* sequences integrated into a chromosome in a region of heterochromatin. (a) When an expression vector for the *lac* repressor was transfected into these cells, *lac* repressors bound to the *lac operator* sites could be visualized in a region of condensed chromatin using an antibody against the *lac* repressor (red). DNA was visualized by staining with DAPI (blue), revealing the nucleus. (b) When an expression vector for the *lac* repressor fused to an activation domain was transfected into these cells, staining as in (a) revealed that the activation domain causes this region of chromatin to decondense into a thinner chromatin fiber that fills a much larger volume of the nucleus. Bar = 1 μm . [Courtesy of Andrew S. Belmont, 1999, *J. Cell Biol.* **145**:1341]

higher-order chromatin structures. The net result of such chromatin remodeling is to facilitate the binding of transcription factors to DNA in chromatin. Some activation domains have been shown to bind to the SWI/SNF complex, and this binding stimulates *in vitro* transcription from chromatin templates (DNA bound to nucleosomes). Thus the SWI/SNF complex represents another type of co-activator complex. Other multi-protein complexes with similar chromatin-remodeling activities have been identified in yeast, raising the possibility that different chromatin-remodeling complexes may be required by distinct families of activators (see chapter opening figure).

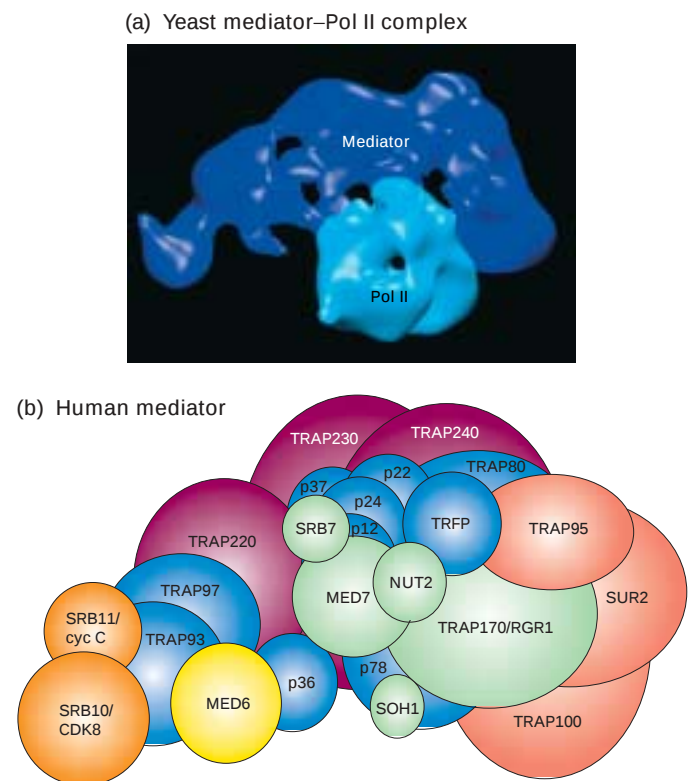
Higher eukaryotes also contain multiprotein complexes with homology to the yeast SWI/SNF complex. These complexes isolated from nuclear extracts of mammalian and *Drosophila* cells have been found to assist binding of transcription factors to their cognate sites in nucleosomal DNA in an ATP-requiring process. The experiment shown in Figure 11-34 dramatically demonstrates how an activation domain can cause decondensation of a region of chromatin. This is thought to result from the interaction of the activation domain with chromatin remodeling and histone acetylase complexes.

Surprisingly, SWI/SNF complexes are also required for the repression of some genes, perhaps because they help expose histone tails to deacetylases or because they assist in the

folding of chromatin into condensed, higher-order structures. Much remains to be learned about how this important class of co-activators and co-repressors alters chromatin structure to influence gene expression.

The Mediator Complex Forms a Molecular Bridge Between Activation Domains and Pol II

Still another type of co-activator, the multiprotein mediator complex, assists more directly in assembly of Pol II preinitiation complexes (Figure 11-35). Some of the ≈ 20 mediator subunits binds to RNA polymerase II, and other mediator subunits bind to activation domains in various activator proteins. Thus mediator can form a molecular bridge between



▲ FIGURE 11-35 Structure of yeast and human mediator complexes. (a) Reconstructed image of mediator from *S. cerevisiae* bound to Pol II. Multiple electron microscopy images were aligned and computer-processed to produce this average image in which the three-dimensional Pol II structure (light blue) is shown associated with the yeast mediator complex (dark blue). (b) Diagrammatic representation of mediator subunits from human cells. Subunits shown in the same color are thought to form a module. Subunits in orange, yellow, and green are homologous with subunits in the yeast mediator complex. Genetic studies in yeast show that mutations in one of the subunits in a module inhibit the association of other subunits in the same module with the rest of the complex. [Part (a) courtesy of Francisco J. Asturias, 2002, *Mol. Cell* **10**:409. Part (b) adapted from S. Malik and R. G. Roeder, 2000, *Trends Biochem. Sci.* **25**:277]

an activator bound to its cognate site in DNA and Pol II at a promoter. In addition, one of the mediator subunits has histone acetylase activity and may function to maintain a promoter region in a hyperacetylated state.

Experiments with temperature-sensitive yeast mutants indicate that some mediator subunits are required for transcription of virtually all yeast genes. These subunits most likely help maintain the overall structure of the mediator complex or bind to Pol II and therefore are required for activation by all activators. In contrast, other mediator subunits are required for activation of specific subsets of genes. DNA microarray analysis of gene expression in mutants with defects in these mediator subunits indicates that each such subunit influences transcription of ≈ 3 –10 percent of all genes (see Figure 9-35 for DNA microarray technique). These mediator subunits are thought to interact with specific activation domains; thus when one subunit is defective, transcription of genes regulated by activators that bind to that subunit is severely depressed but transcription of other genes is unaffected. Consistent with this explanation are binding studies showing that some activation domains do indeed interact with specific mediator subunits.

Large mediator complexes, isolated from cultured mammalian cells, are required for mammalian activators to stimulate transcription by Pol II in vitro. Since genes encoding homologs of the mammalian mediator subunits have been identified in the genome sequences of *C. elegans* and *Drosophila*, it appears that most multicellular animals (metazoans) have homologous mediator complexes. About one third of the metazoan mediator subunits are clearly homologous to yeast mediator subunits (see Figure 11-35b). But the remaining subunits, which appear to be distinct from any yeast proteins, may interact with activation domains that

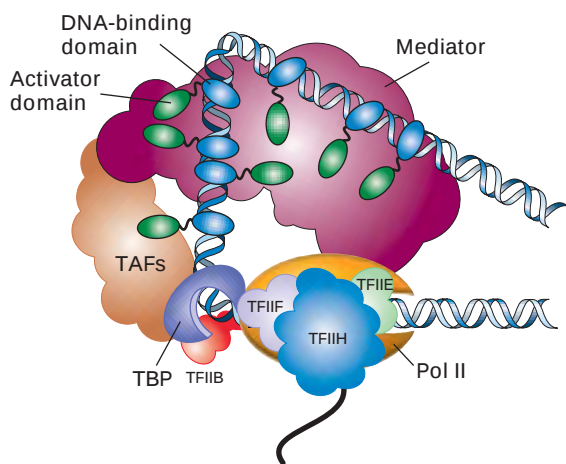
are not found in yeast. As with yeast mediator, some of the mammalian mediator subunits have been shown to interact with specific activation domains. For example, the Sur2 subunit of mammalian mediator binds to the activation domain of a TCF transcription factor that controls expression of the *EGR-1* gene (see Figure 11-19). The function of this TCF activator in vivo normally is regulated in response to specific protein hormones present in serum. Mouse embryonic stem cells with a knockout of the *sur2* gene fail to induce expression of EGR-1 protein in response to serum, whereas multiple other activators function normally in the mutant cells. This finding implicates the mediator Sur2 subunit in the activating function of TCF.

The various experimental results indicating that individual mediator subunits bind to specific activation domains suggest that multiple activators influence transcription from a single promoter by interacting with a mediator complex simultaneously (Figure 11-36). Activators bound at enhancers or promoter-proximal elements can interact with mediator associated with a promoter because DNA is flexible and can form a loop bringing the regulatory regions and the promoter close together. Such loops have been observed in experiments with the *E. coli* NtrC activator and σ^{54} -RNA polymerase (see Figure 4-17). The multiprotein nucleoprotein complexes that form on eukaryotic promoters may comprise as many as 100 polypeptides with a total mass of ≈ 3 megadaltons (MDa), as large as a ribosome.

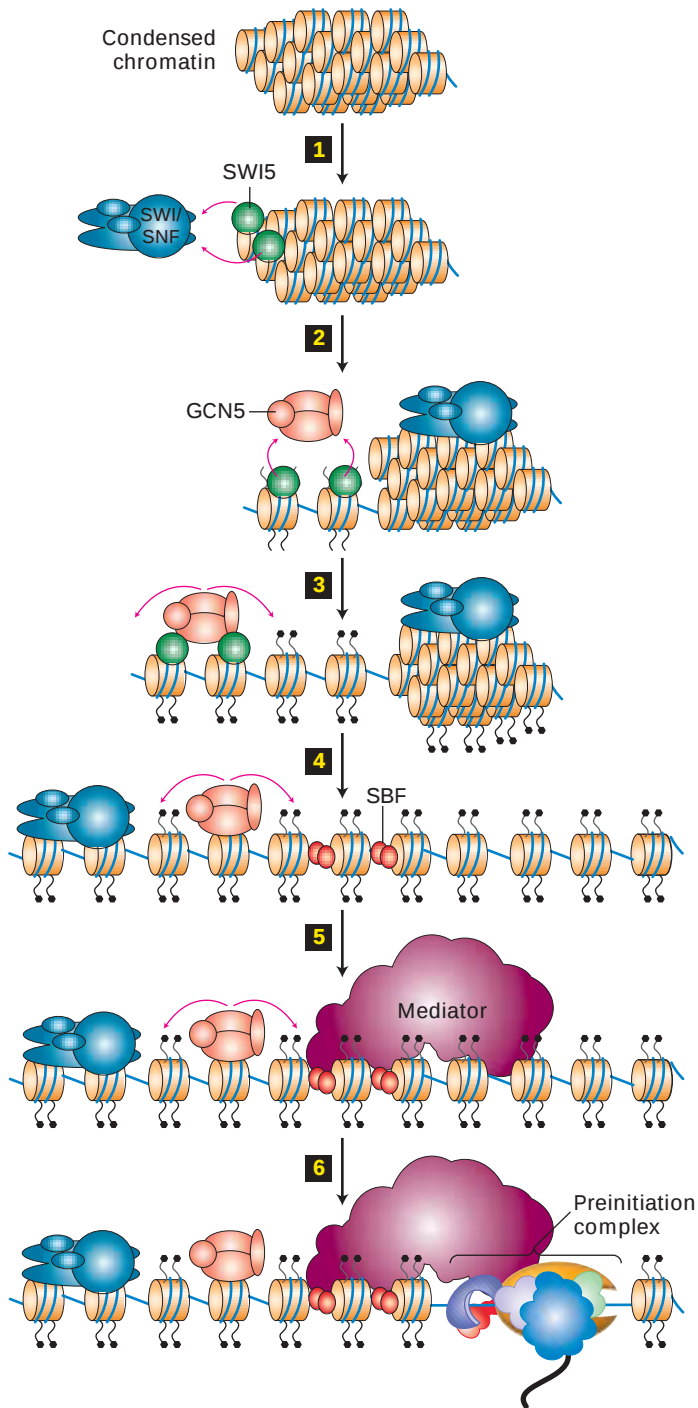
Transcription of Many Genes Requires Ordered Binding of Activators and Action of Co-Activators

We can now extend the model of Pol II transcription initiation in Figure 11-27 to take into account the role of activators and co-activators. These accessory proteins function not only to make genes within nucleosomal DNA accessible to general transcription factors and Pol II but also directly recruit Pol II to promoter regions.

Recent studies have analyzed the order in which activators bind to a transcription-control region and interact with co-activators as a gene is induced. Such studies show that assembly of preinitiation complexes depends on multiple protein-DNA and protein-protein interactions, as illustrated in Figure 11-37 depicting activation of the yeast *HO* gene. This gene encodes a sequence-specific nuclease that initiates mating-type switching in haploid yeast cells (see Figure 11-28). Activation of the *HO* gene begins with binding of the SWI5 activator to an upstream enhancer. Bound SWI5 then interacts with the SWI/SNF chromatin-remodeling complex and GCN5-containing histone acetylase complex. Once the chromatin in the *HO* control region is decondensed and hyperacetylated, a second activator, SBF, can bind to several sites in the promoter-proximal region. Subsequent binding of the mediator complex by SBF then leads to assembly of the transcription preinitiation complex containing Pol II and the general transcription factors shown in Figure 11-36.



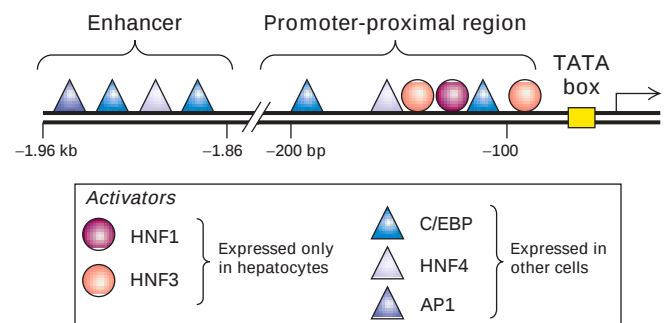
▲ **FIGURE 11-36 Model of several DNA-bound activators interacting with a single mediator complex.** The ability of different mediator subunits to interact with specific activation domains may contribute to the integration of signals from several activators at a single promoter. See the text for discussion.



We can now see that the assembly of a preinitiation complex and stimulation of transcription at a promoter results from the interaction of several activators with various multi-protein co-activator complexes. These include chromatin-remodeling complexes, histone acetylase complexes, and a mediator complex. While much remains to be learned about these processes, it is clear that the net result of these multiple molecular events is that activation of transcription at a promoter depends on highly cooperative interactions initi-

◀ **FIGURE 11-37 Ordered binding and interaction of activators and co-activators leading to transcription of the yeast *HO* gene.** Step **1**: Initially, the *HO* gene is packaged into condensed chromatin. Activation begins when the SWI5 activator binds to enhancer sites 1200–1400 base pairs upstream of the start site and interacts with the SWI/SNF chromatin-remodeling complex. Step **2**: The SWI/SNF complex acts to decondense the chromatin, thereby exposing histone tails. Step **3**: A GCN5-containing histone acetylase complex associates with bound SWI5 and acetylates histone tails in the *HO* locus as SWI/SNF continues to decondense adjacent chromatin. Step **4**: SWI5 is released from the DNA, but the SWI/SNF and GCN5 complexes remain associated with the *HO* control region (in the case of GCN5, by poorly understood interactions). Their action allows the SBF activator to bind several sites in the promoter-proximal region. Step **5**: SBF then binds the mediator complex. Step **6**: Subsequent binding of Pol II and general transcription factors results in assembly of a transcription preinitiation complex whose components are detailed in Figure 11-37. [Adapted from C. J. Fry and C. L. Peterson, 2001, *Curr. Biol.* **11**:R185. See also M. P. Cosma et al., 1999, *Cell* **97**:299, and M. P. Cosma et al., 2001, *Mol. Cell* **7**:1213.]

ated by several activators. This allows genes to be regulated in a cell-type-specific manner by specific combinations of transcription factors. The *TTR* gene, which encodes transthyretin in mammals, is a good example of this. As noted earlier, transthyretin is expressed in hepatocytes and in choroid plexus cells. Transcription of the *TTR* gene in hepatocytes is controlled by at least five different transcriptional activators (Figure 11-38). Even though three of these activators—HNF4, C/EBP, and AP1—are also expressed in cells of the intestine and kidney, *TTR* transcription does not occur in these cells, because all five activators are required but HNF1 and HNF3 are missing. Other hepatocyte-



▲ **FIGURE 11-38 Transcription-control region of the mouse transthyretin (*TTR*) gene.** Binding sites for the five activators required for transcription of *TTR* in hepatocytes are indicated. The complete set of activators is expressed at the required concentrations to stimulate transcription only in hepatocytes. A different set of activators stimulates transcription in choroid plexus cells. [See R. Costa et al., 1989, *Mol. Cell Biol.* **9**:1415, and K. Xanthopoulos et al., 1989, *Proc. Nat'l. Acad. Sci. USA* **86**:4117.]

specific enhancers and promoter-proximal regions that regulate additional genes expressed only in hepatocytes contain binding sites for other specific combinations of transcription factors found only in these cells, together with those expressed ubiquitously.

The Yeast Two-Hybrid System Exploits Activator Flexibility to Detect cDNAs That Encode Interacting Proteins

A powerful molecular genetic method called the **yeast two-hybrid system** exploits the flexibility in activator structures to identify genes whose products bind to a specific protein of interest. Because of the importance of protein-protein interactions in virtually every biological process, the yeast two-hybrid system is used widely in biological research.

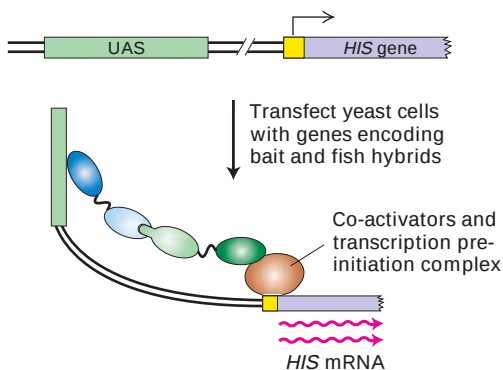
This method employs a yeast vector for expressing a DNA-binding domain and flexible linker region without the associated activation domain, such as the deleted GAL4 containing amino acids 1–692 (see Figure 11-17). A cDNA sequence encoding a protein or protein domain of interest, called the *bait domain*, is fused in frame to the flexible linker region so that the vector will express a hybrid protein composed of the DNA-binding domain, linker region, and bait domain (Figure 11-39a, left). A cDNA library is cloned into multiple copies of a second yeast vector that encodes a strong activation domain and flexible linker, to produce a vector library expressing multiple hybrid proteins, each containing a different *fish domain* (Figure 11-39a, right).

The bait vector and library of fish vectors are then transfected into engineered yeast cells in which the only copy of

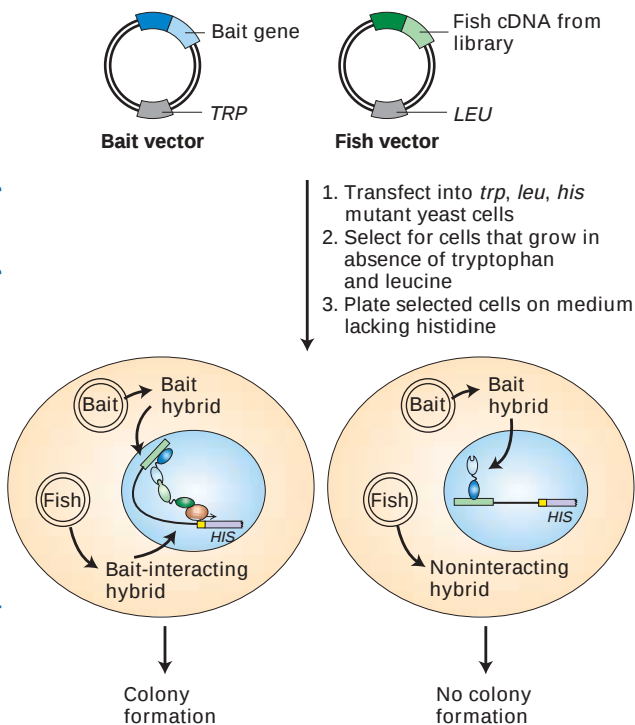
(a) Hybrid proteins



(b) Transcriptional activation by hybrid proteins in yeast



(c) Fishing for proteins that interact with bait domain



◀ EXPERIMENTAL FIGURE 11-39 The yeast two-hybrid system provides a way of screening a cDNA library for clones encoding proteins that interact with a specific protein of interest.

This is a common technique for screening a cDNA library for clones encoding proteins that interact with a specific protein of interest. (a) Two vectors are constructed containing genes that encode hybrid (chimeric) proteins. In one vector (*left*), coding sequence for the DNA-binding domain of a transcription factor is fused to the sequences for a known protein, referred to as the “bait” domain (light blue). The second vector (*right*) expresses an activation domain fused to a “fish” domain (green) that interacts with the bait domain. (b) If yeast cells are transformed with vectors expressing both hybrids, the bait and fish portions of the chimeric proteins interact to produce a functional transcriptional activator. In this example, the activator promotes transcription of a *HIS* gene. One end of this protein complex binds to the upstream activating sequence (UAS) of the *HIS3* gene; the other end, consisting of the activation domain, stimulates assembly of the transcription preinitiation complex (orange) at the promoter (yellow). (c) To screen a cDNA library for clones encoding proteins that interact with a particular bait protein of interest, the library is cloned into the vector encoding the activation domain so that hybrid proteins are expressed. The bait vector and fish vectors contain wild-type selectable genes (e.g., a *TRP* or *LEU* gene). The only transformed cells that survive the indicated selection scheme are those that express the bait hybrid and a fish hybrid that interacts with it. See the text for discussion. [See S. Fields and O. Song, 1989, *Nature* **340**:245.]

a gene required for histidine synthesis (*HIS*) is under control of a UAS with binding sites for the DNA-binding domain of the hybrid bait protein. Transcription of the *HIS* gene requires activation by proteins bound to the UAS. Transformed cells that express the bait hybrid and an *interacting* fish hybrid will be able to activate transcription of the *HIS* gene (Figure 11-39b). This system works because of the flexibility in the spacing between the DNA-binding and activation domains of eukaryotic activators.

A two-step selection process is used (Figure 11-39c). The bait vector also expresses a wild-type *TRP* gene, and the hybrid vector expresses a wild-type *LEU* gene. Transfected cells are first grown in a medium that lacks tryptophan and leucine but contains histidine. Only cells that have taken up the bait vector and one of the fish plasmids will survive in this medium. The cells that survive then are plated on a medium that lacks histidine. Those cells expressing a fish hybrid that does not bind to the bait hybrid cannot transcribe the *HIS* gene and consequently will not form a colony on medium lacking histidine. The few cells that express a bait-binding fish hybrid will grow and form colonies in the absence of histidine. Recovery of the fish vectors from these colonies yields cDNAs encoding protein domains that interact with the bait domain.

KEY CONCEPTS OF SECTION 11.5

Molecular Mechanisms of Transcription Activation and Repression

- Eukaryotic transcription activators and repressors exert their effects largely by binding to multisubunit co-activators or co-repressors that influence assembly of Pol II transcription preinitiation complexes either by modulating chromatin structure (indirect effect) or by interacting with Pol II and general transcription factors (direct effect).
- The DNA in condensed regions of chromatin (heterochromatin) is relatively inaccessible to transcription factors and other proteins, so that gene expression is repressed.
- The interactions of several proteins with each other and with the hypoacetylated N-terminal tails of histones H3 and H4 are responsible for the chromatin-mediated repression of transcription that occurs in the telomeres and the silent mating-type loci in *S. cerevisiae*. (see Figure 11-30).
- Some repression domains function by interacting with co-repressors that are histone deacetylase complexes. The subsequent deacetylation of histone N-terminal tails in nucleosomes near the repressor-binding site inhibits interaction between the promoter DNA and general transcription factors, thereby repressing transcription initiation (see Figure 11-32a).
- Some activation domains function by binding multiprotein co-activator complexes such as histone acetylase

complexes. The subsequent hyperacetylation of histone N-terminal tails in nucleosomes near the activator-binding site facilitates interactions between the promoter DNA and general transcription factors, thereby stimulating transcription initiation (see Figure 11-32b).

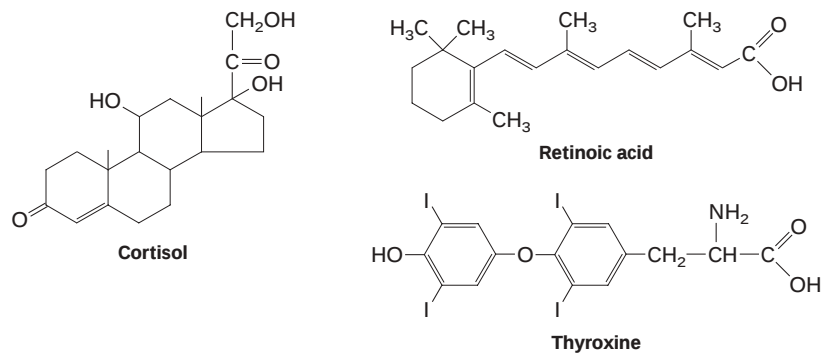
- SWI/SNF chromatin-remodeling factors constitute another type of co-activator. These multisubunit complexes can transiently dissociate DNA from histone cores in an ATP-dependent reaction and may also decondense regions of chromatin, thereby promoting the binding of DNA-binding proteins needed for initiation to occur at some promoters.
- Mediator, another type of co-activator, is an ≈ 20 -subunit complex that forms a molecular bridge between activation domains and RNA polymerase II by binding directly to the polymerase and activation domains. By binding to several different activators simultaneously, mediator probably helps integrate the effects of multiple activators on a single promoter (see Figure 11-36).
- Activators bound to a distant enhancer can interact with transcription factors bound to a promoter because DNA is flexible and the intervening DNA can form a large loop.
- The highly cooperative assembly of preinitiation complexes *in vivo* generally requires several activators. A cell must produce the specific set of activators required for transcription of a particular gene in order to express that gene.
- The yeast two-hybrid system is widely used to detect cDNAs encoding protein domains that bind to a specific protein of interest (see Figure 11-39).

11.6 Regulation of Transcription-Factor Activity

We have seen in the preceding discussion how combinations of activators and repressors that bind to specific DNA regulatory sequences control transcription of eukaryotic genes. Whether or not a specific gene in a multicellular organism is expressed in a particular cell at a particular time is largely a consequence of the concentrations and activities of the transcription factors that interact with the regulatory sequences of that gene. Which transcription factors are expressed in a particular cell type, and the amounts produced, is determined by multiple regulatory interactions between transcription-factor genes that occur during the development and differentiation of a particular cell type. In Chapters 15 and 22, we present examples of such regulatory interactions during development and discuss the principles of development and differentiation that have emerged from these examples.

Not only is the expression of transcription factors by a cell regulated, but the activities of those factors expressed in a particular cell type commonly are further controlled

► **FIGURE 11-40 Examples of hormones that bind to nuclear receptors.** These and related lipid-soluble hormones bind to receptors located in the cytosol or nucleus. The ligand-receptor complex functions as a transcription activator.



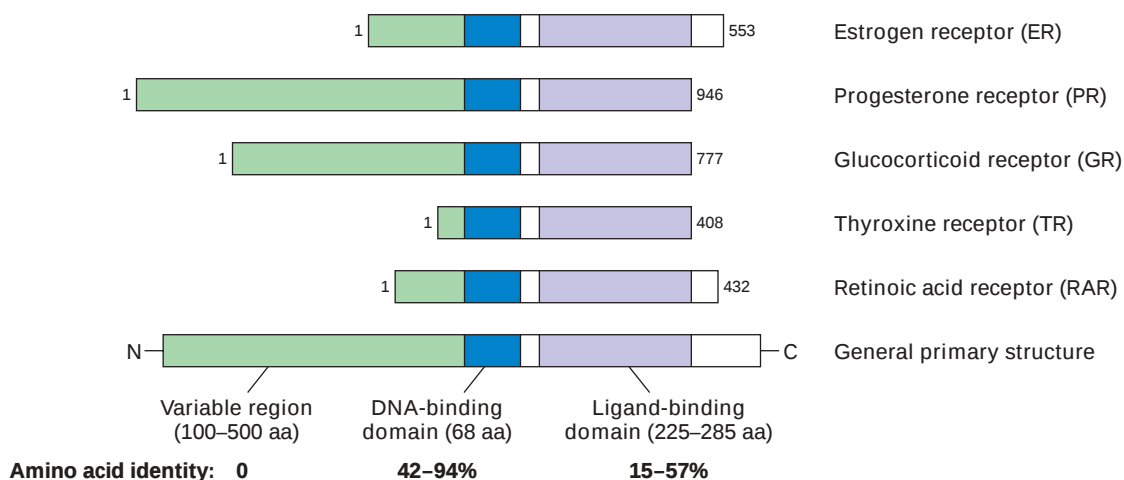
indirectly as the result of interactions between proteins on the surfaces of neighboring cells and by extracellular hormones and growth factors. In multicellular organisms, these latter signaling molecules are secreted from one cell type and affect the function of cells that may be nearby or at a different location in the organism. One major group of extracellular signals comprises peptides and proteins, which bind to receptors in the plasma membrane. Ligand binding to these receptors triggers intracellular **signal-transduction** pathways. In Chapters 14 and 15, we describe the major types of cell-surface receptors and intracellular signaling pathways that regulate transcription-factor activity.

In this section, we discuss the second major group of extracellular signals, the small, lipid-soluble hormones—including many different steroid hormones, retinoids, and thyroid hormones—that can diffuse through plasma and nuclear membranes and interact directly with the transcription factors they control (Figure 11-40). As noted earlier, the intracellular receptors for most of these lipid-soluble hormones,

which constitute the *nuclear-receptor superfamily*, function as transcription activators when bound to their ligands.

All Nuclear Receptors Share a Common Domain Structure

Cloning and sequencing of the genes encoding various nuclear receptors revealed a remarkable conservation in their amino acid sequences and three functional regions (Figure 11-41). All the nuclear receptors have a unique N-terminal region of variable length (100–500 amino acids). Portions of this variable region function as activation domains in some nuclear receptors. The DNA-binding domain maps near the center of the primary sequence and has a repeat of the C₄ zinc-finger motif. The hormone-binding domain, located near the C-terminal end, contains a hormone-dependent activation domain. In some nuclear receptors, the hormone-binding domain functions as a repression domain in the absence of ligand.



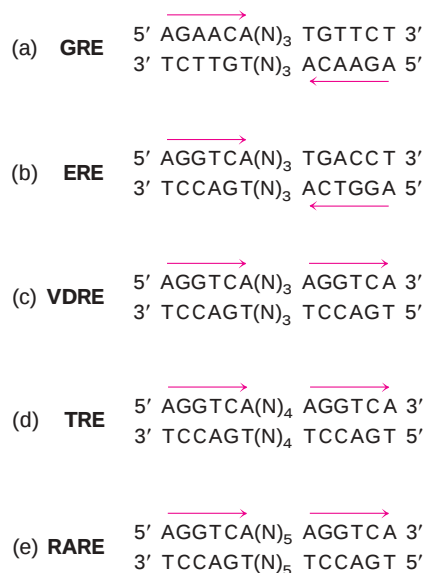
▲ **FIGURE 11-41 General design of transcription factors in nuclear-receptor superfamily.** The centrally located DNA-binding domain exhibits considerable sequence homology among different receptors and contains two copies of the C₄ zinc-finger

motif. The C-terminal hormone-binding domain exhibits somewhat less homology. The N-terminal regions in various receptors vary in length, have unique sequences, and may contain one or more activation domains. [See R. M. Evans, 1988, *Science* **240**:889.]

Nuclear-Receptor Response Elements Contain Inverted or Direct Repeats

The characteristic nucleotide sequences of the DNA sites, called *response elements*, that bind several nuclear receptors have been determined. The sequences of the consensus response elements for the glucocorticoid and estrogen receptors are 6-bp inverted repeats separated by any three base pairs (Figure 11-42a, b). This finding suggested that the cognate steroid hormone receptors would bind to DNA as symmetrical dimers, as was later shown from the x-ray crystallographic analysis of the homodimeric glucocorticoid receptor's C₄ zinc-finger DNA-binding domain (see Figure 11-21b).

Some nuclear-receptor response elements, such as those for the receptors that bind vitamin D₃, thyroid hormone, and retinoic acid, are direct repeats of the same sequence recognized by the estrogen receptor, separated by three to five base pairs (Figure 11-42c–e). The specificity for responding to these different hormones by binding distinct receptors is determined by the spacing between the repeats. The receptors that bind to such direct-repeat response elements do so as heterodimers with a common nuclear-receptor monomer called RXR. The vitamin D₃ response element, for example, is bound by the RXR-VDR heterodimer,



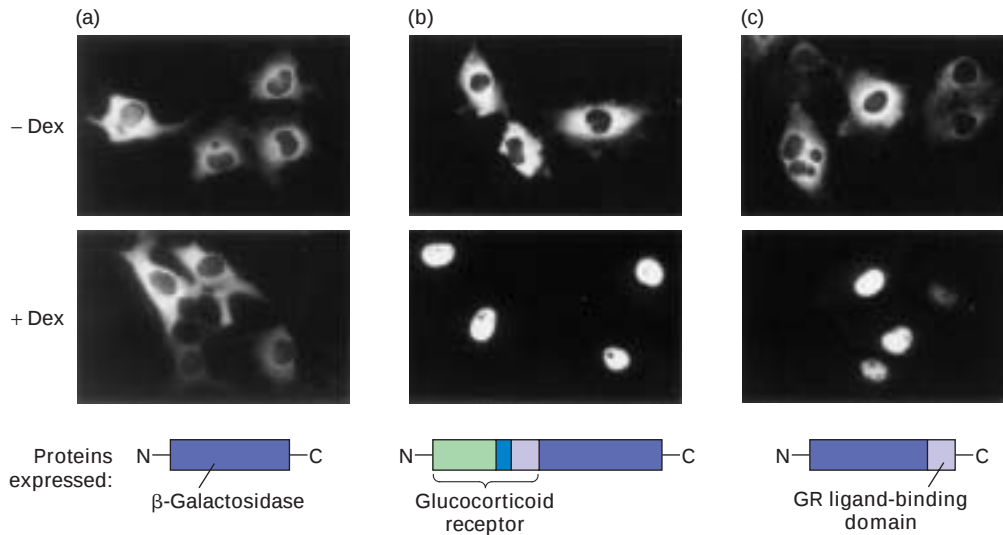
▲ **FIGURE 11-42 Consensus sequences of DNA response elements that bind three nuclear receptors.** The response elements for the glucocorticoid receptor (GRE) and estrogen receptor (ERE) contain inverted repeats that bind these homodimeric proteins. The response elements for heterodimeric receptors contain a common direct repeat separated by three to five base pairs, for the vitamin D₃ receptor (VDRE), thyroid hormone receptor (TRE), and retinoic acid receptor (RARE). The repeat sequences are indicated by red arrows. [See K. Umesono et al., 1991, *Cell* **65**:1255, and A. M. Naar et al., 1991, *Cell* **65**:1267]

and the retinoic acid response element is bound by RXR-RAR. The monomers composing these heterodimers interact with each other in such a way that the two DNA-binding domains lie in the same rather than inverted orientation, allowing the RXR heterodimers to bind to direct repeats of the binding site for each monomer. In contrast, the monomers in homodimeric nuclear receptors (e.g., GRE and ERE) have an inverted orientation.

Hormone Binding to a Nuclear Receptor Regulates Its Activity as a Transcription Factor

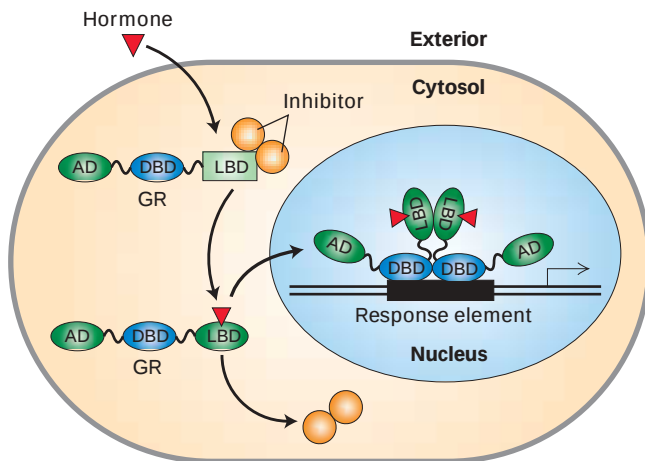
The mechanism whereby hormone binding controls the activity of nuclear receptors differs for heterodimeric and homodimeric receptors. Heterodimeric nuclear receptors (e.g., RXR-VDR, RXR-TR, and RXR-RAR) are located exclusively in the nucleus. In the absence of their hormone ligand, they repress transcription when bound to their cognate sites in DNA. They do so by directing histone deacetylation at nearby nucleosomes by the mechanism described earlier (see Figure 11-32a). As we saw earlier, in the presence of hormone, the ligand-binding domain of the RAR monomer undergoes a dramatic conformational change compared with the ligand-binding domain of a nuclear receptor in the absence of hormone (see Figure 11-25). In the ligand-bound conformation, heterodimeric nuclear receptors containing RXR can direct hyperacetylation of histones in nearby nucleosomes, thereby reversing the repressing effects of the free ligand-binding domain. In the presence of ligand, ligand-binding domains of nuclear receptors also bind mediator, stimulating preinitiation complex assembly.

In contrast to heterodimeric nuclear receptors, homodimeric receptors are found in the cytoplasm in the absence of their ligands. Hormone binding to these receptors leads to their translocation to the nucleus. The hormone-dependent translocation of the homodimeric glucocorticoid receptor (GR) was demonstrated in the transfection experiments shown in Figure 11-43. The GR hormone-binding domain alone mediates this transport. Subsequent studies showed that, in the absence of hormone, GR is anchored in the cytoplasm as a large protein aggregate complexed with inhibitor proteins, including Hsp90, a protein related to Hsp70, the major heat-shock chaperone in eukaryotic cells. As long as the receptor is confined to the cytoplasm, it cannot interact with target genes and hence cannot activate transcription. Hormone binding to a homodimeric nuclear receptor releases the inhibitor proteins, allowing the receptor to enter the nucleus, where it can bind to response elements associated with target genes (Figure 11-44). Once the receptor with bound hormone binds to a response element, it activates transcription by interacting with chromatin-remodeling and histone acetylase complexes and mediator.



▲ EXPERIMENTAL FIGURE 11-43 Fusion proteins from expression vectors demonstrate that the hormone-binding domain of the glucocorticoid receptor (GR) mediates translocation to the nucleus in the presence of hormone. Cultured animal cells were transfected with expression vectors encoding the proteins diagrammed at the bottom. Immunofluorescence with a labeled antibody specific for β -galactosidase was used to detect the expressed proteins in transfected cells. (a) In cells that expressed β -galactosidase alone, the enzyme

was localized to the cytoplasm in the presence and absence of the glucocorticoid hormone dexamethasone (Dex). (b) In cells that expressed a fusion protein consisting of β -galactosidase and the entire glucocorticoid receptor (GR), the fusion protein was present in the cytoplasm in the absence of hormone but was transported to the nucleus in the presence of hormone. (c) Cells that expressed a fusion protein composed of β -galactosidase and just the GR ligand-binding domain (light purple) also exhibited hormone-dependent transport of the fusion protein to the nucleus. [From D. Picard and K. R. Yamamoto, 1987, *EMBO J.* 6:3333; courtesy of the authors.]



▲ FIGURE 11-44 Model of hormone-dependent gene activation by a homodimeric nuclear receptor. In the absence of hormone, the receptor is kept in the cytoplasm by interaction between its ligand-binding domain (LBD) and inhibitor proteins. When hormone is present, it diffuses through the plasma membrane and binds to the ligand-binding domain, causing a conformational change that releases the receptor from the inhibitor proteins. The receptor with bound ligand is then translocated into the nucleus, where its DNA-binding domain (DBD) binds to response elements, allowing the ligand-binding domain and an additional activation domain (AD) at the N-terminus to stimulate transcription of target genes.

KEY CONCEPTS OF SECTION 11.6

Regulation of Transcription-Factor Activity

- The activities of many transcription factors are indirectly regulated by binding of extracellular proteins and peptides to cell-surface receptors. These receptors activate intracellular signal transduction pathways that regulate specific transcription factors through a variety of mechanisms discussed in Chapters 13 and 14.
- Nuclear receptors constitute a superfamily of dimeric C_4 zinc-finger transcription factors that bind lipid-soluble hormones and interact with specific response elements in DNA (see Figure 11-41).
- Hormone binding to nuclear receptors induces conformational changes that modify their interactions with other proteins.
- Heterodimeric nuclear receptors (e.g., those for retinoids, vitamin D, and thyroid hormone) are found only in the nucleus. In the absence of hormone, they repress transcription of target genes with the corresponding response element. When bound to their ligands, they activate transcription.
- Steroid hormone receptors are homodimeric nuclear receptors. In the absence of hormone, they are trapped in the cytoplasm by inhibitor proteins. When bound to their ligands, they can translocate to the nucleus and activate transcription of target genes (see Figure 11-44).

11.7 Regulated Elongation and Termination of Transcription

In eukaryotes, the mechanisms for terminating transcription differ for each of the three RNA polymerases. Transcription of pre-rRNA genes by RNA polymerase I is terminated by a mechanism that requires a polymerase-specific termination factor. This DNA-binding protein binds to a specific DNA sequence downstream of the transcription unit. Efficient termination requires that the termination factor bind to the template DNA in the correct orientation. Purified RNA polymerase III terminates after polymerizing a series of U residues. The deoxy(A)_n-ribo(U)_n DNA-RNA hybrid that results when a stretch of U's are synthesized is particularly unstable compared with all other base-paired sequences. The ease with which this hybrid can be melted probably contributes to the mechanism of termination by RNA polymerase III.

In most mammalian protein-coding genes transcribed by RNA polymerase II, once the polymerase has transcribed beyond about fifty bases, further elongation is highly processive and does not terminate until after a sequence is transcribed that directs cleavage and polyadenylation of the RNA at the sequence that forms the 3' end of the encoded mRNA. RNA polymerase II then can terminate at multiple sites located over a distance of 0.5–2 kb beyond this poly(A) addition site. Experiments with mutant genes show that termination is coupled to the process that cleaves and polyadenylates the 3' end of a transcript, which is discussed in the next chapter. Biochemical and chromatin immunoprecipitation experiments suggest that the protein complex that cleaves and polyadenylates the nascent mRNA transcript at specific sequences associates with the phosphorylated carboxyl-terminal domain (CTD) of RNA polymerase II following initiation (see Figure 11-27). This cleavage/polyadenylation complex may suppress termination by RNA polymerase II until the sequence signaling cleavage and polyadenylation is transcribed by the polymerase.

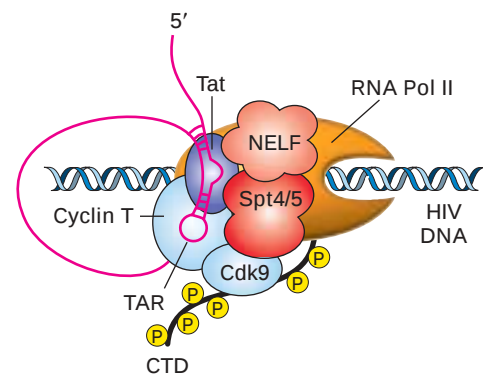
While transcription termination is unregulated for most genes, for some specific genes, a choice is made between elongation and termination or pausing within a few tens of bases from the transcription start site. This choice between elongation and termination or pausing can be regulated; thus expression of the encoded protein is controlled not only by transcription initiation, but also by control of transcription elongation early in the transcription unit. We discuss two examples of such regulation next.

Transcription of the HIV Genome Is Regulated by an Antitermination Mechanism

Currently, transcription of the human immunodeficiency virus (HIV) genome by RNA polymerase II provides the best-understood example of regulated transcription termination in eukaryotes. Efficient expression of HIV genes requires a small viral protein encoded at the *tat* locus. Cells infected

with *tat*[−] mutants produce short viral transcripts that hybridize to restriction fragments containing promoter-proximal regions of the HIV DNA but not to restriction fragments farther downstream from the promoter. In contrast, cells infected with wild-type HIV synthesize long viral transcripts that hybridize to restriction fragments throughout the single HIV transcription unit. Thus Tat protein functions as an *antitermination factor*, permitting RNA polymerase II to read through a transcriptional block. Since antitermination by Tat protein is required for HIV replication, further understanding of this gene-control mechanism may offer possibilities for designing effective therapies for acquired immunodeficiency syndrome (AIDS).

Tat is a sequence-specific RNA-binding protein. It binds to the RNA copy of a sequence called TAR, which is located near the 5' end of the HIV transcript. The TAR sequence folds into an RNA hairpin with a bulge in the middle of the stem (Figure 11-45). TAR contains two binding sites: one that interacts with Tat and one that interacts with a cellular protein called *cyclin T*. As depicted in Figure 11-45, the HIV Tat protein and cellular cyclin T each bind to TAR RNA and also interact directly with each other so that they bind cooperatively, much like the cooperative binding of DNA-binding transcription factors (see Figure 11-24). Interaction of cyclin T with a protein kinase called CDK9 activates the kinase, whose substrate is the CTD of RNA polymerase II. In vitro transcription studies using a specific inhibitor of CDK9 suggest that RNA polymerase II molecules that initiate transcription on the HIV promoter terminate after transcribing ≈50 bases unless the CTD is hyperphosphorylated by CDK9. Cooperative binding of cyclin T and Tat to the TAR sequence at the 5' end of the HIV transcript positions CDK9 so that it can phosphorylate the CTD, thereby preventing termination and permitting the polymerase to continue chain elongation.



▲ FIGURE 11-45 Model of antitermination complex composed of HIV Tat protein and several cellular proteins.

The TAR element in the HIV transcript contains sequences recognized by Tat and the cellular protein cyclin T. Cyclin T activates and helps position the protein kinase CDK9 near its substrate, the CTD of RNA polymerase II. See the text for a more detailed discussion. [See P. Wei et al., 1998, *Cell* **92**:451; T. Wada et al., 1998, *Genes & Devel.* **12**:357; and Y. Yamaguchi et al., 1999, *Cell* **97**:41]

Several additional cellular proteins, including Spt4 and Spt5 and the NELF complex, participate in the process by which HIV Tat controls elongation versus termination (see Figure 11-45). Experiments with the specific inhibitor of CDK9 mentioned above and with *spt4* and *spt5* yeast mutants indicate that these cellular proteins are required for transcription elongation beyond ≈ 50 bases for most cellular genes. But for most genes, these proteins appear to function constitutively, that is, without being regulated. As discussed in Chapter 12, RNA polymerase II pausing instigated by Spt4/5 and NELF is thought to delay elongation until mRNA processing factors associate with the phosphorylated CTD. Further phosphorylation of the CTD by cyclin T–CDK9 (also known as pTEFb) appears to reverse this pause and allow elongation to continue. Currently, it is not clear why this process is not constitutive for the HIV promoter, where cooperative binding of HIV Tat and cyclin T to the TAR RNA sequence is required for efficient elongation.

Promoter-Proximal Pausing of RNA Polymerase II Occurs in Some Rapidly Induced Genes

The *heat-shock genes* (e.g., *hsp70*) illustrate another mechanism for regulating RNA chain elongation in eukaryotes. During transcription of these genes, RNA polymerase II pauses after transcribing ≈ 25 nucleotides but does not terminate transcription (as it does when transcribing the HIV genome in the absence of Tat protein). The paused polymerase remains associated with the nascent RNA and template DNA, until conditions occur that lead to activation of HSTF (heat-shock transcription factor). Subsequent binding of activated HSTF to specific sites in the promoter-proximal region of heat-shock genes stimulates the paused polymerase to continue chain elongation and promotes rapid reinitiation by additional RNA polymerase II molecules.

The pausing during transcription of heat-shock genes initially was discovered in *Drosophila*, but a similar mechanism most likely occurs in other eukaryotes. Heat-shock genes are induced by intracellular conditions that denature proteins (such as elevated temperature, “heat shock”). Some encode proteins that are relatively resistant to denaturing conditions and act to protect other proteins from denaturation; others are chaperonins that refold denatured proteins (see Chapter 3). The mechanism of transcriptional control that evolved to regulate expression of these genes permits a rapid response: these genes are already paused in a state of suspended transcription and therefore, when an emergency arises, require no time to remodel and acetylate chromatin over the promoter and assemble a transcription preinitiation complex.

terminated until an RNA sequence is synthesized that specifies a site of cleavage and polyadenylation.

- Transcription of the HIV genome by RNA polymerase II is regulated by an antitermination mechanism that requires cooperative binding by the virus-encoded Tat protein and cyclin T to the TAR sequence near the 5' end of the HIV RNA.

- During transcription of *Drosophila* heat-shock genes, RNA polymerase II pauses within the downstream promoter-proximal region; this interruption in transcription is released when the HSTF transcription factor is activated, resulting in very rapid transcription of the heat-shock genes in response to the accumulation of denatured proteins.

11.8 Other Eukaryotic Transcription Systems

We conclude this chapter with a brief discussion of transcription initiation by the other two eukaryotic nuclear RNA polymerases, Pol I and Pol III, and by the distinct polymerases that transcribe mitochondrial and chloroplast DNA. Although these systems, particularly their regulation, are less thoroughly understood than transcription by RNA polymerase II, they are also fundamental to the life of eukaryotic cells.

Transcription Initiation by Pol I and Pol III Is Analogous to That by Pol II

The formation of transcription-initiation complexes involving Pol I and Pol III is similar in some respects to assembly of Pol II initiation complexes (see Figure 11-27). However, each of the three eukaryotic nuclear RNA polymerases requires its own polymerase-specific general transcription factors and recognizes different DNA control elements. Moreover, neither Pol I nor Pol III requires ATP hydrolysis to initiate transcription, whereas Pol II does.

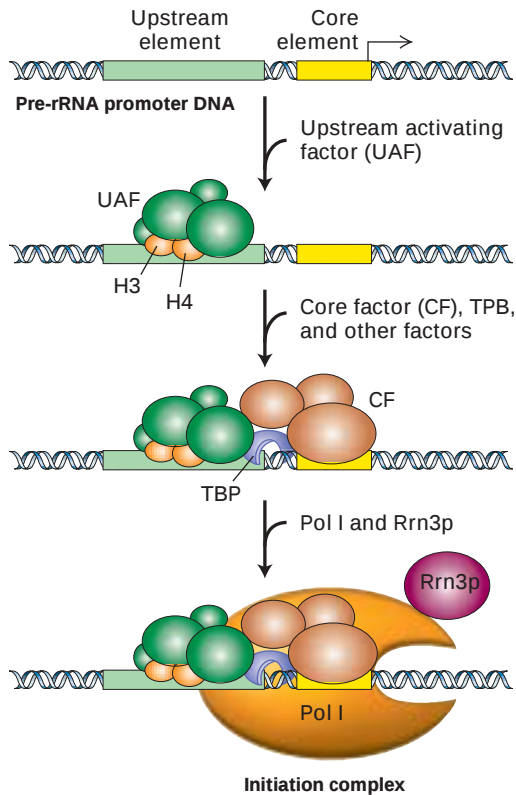
Transcription initiation by Pol I, which synthesizes pre-rRNA, and by Pol III, which synthesizes tRNAs, 5S rRNA, and other short, stable RNAs, has been characterized most extensively in *S. cerevisiae* using both biochemical and genetic approaches. It is clear that synthesis of tRNAs and of rRNAs, which are incorporated into ribosomes, is tightly coupled to the rate of cell growth and proliferation. However, much remains to be learned about how transcription initiation by Pol I and Pol III is regulated so that synthesis of pre-rRNA, 5S rRNA, and tRNAs is coordinated with the growth and replication of cells.

Initiation by Pol I The regulatory elements directing Pol I initiation are similarly located relative to the transcription start site in both yeast and mammals. A *core element* spanning the transcription start site from -40 to $+5$ is essential for Pol I transcription. An additional *upstream element* extending from roughly -155 to -60 stimulates in vitro Pol I transcription tenfold.

KEY CONCEPTS FOR SECTION 11.7

Regulated Elongation and Termination of Transcription

- Different mechanisms of transcription termination are employed by each of the eukaryotic nuclear RNA polymerases. Transcription of most protein-coding genes is not



▲ **FIGURE 11-46 In vitro assembly of the yeast Pol I transcription initiation complex.** UAF and CF, both multimeric general transcription factors, bind to the upstream element (UE) and core element, respectively, in the promoter DNA. TBP and a monomeric factor (Rrn3p) associated with RNA polymerase I (Pol I) also participate in forming the initiation complex. [Adapted from N. Nomura, 1998, in R. M. Paule, *Transcription of Ribosomal RNA Genes by Eukaryotic RNA Polymerase I*, Landes Bioscience, pp. 157–172.]

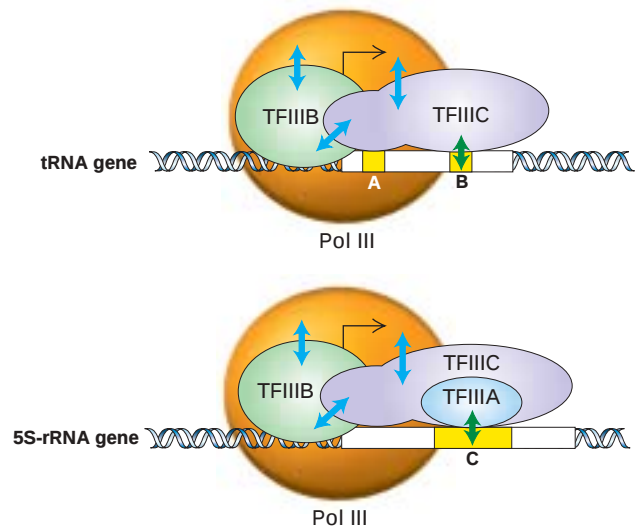
Assembly of a fully active Pol I initiation complex begins with binding of a multimeric upstream activating factor (UAF) to the upstream element (Figure 11-46). Two of the six subunits composing UAF are histones, which probably participate in DNA binding. Next, a trimeric core factor binds to the core element together with TBP, which makes contact with both the bound UAF and the core factor. Finally, a preformed complex of Pol I and Rrn3p associates with the bound proteins, positioning Pol I near the start site. In human cells, TBP is stably bound to three other polypeptides, forming an initiation factor called *SL1* that binds to the core promoter element and is functionally equivalent to yeast core factor plus TBP.

Initiation by Pol III Unlike protein-coding genes and pre-rRNA genes, the promoter regions of tRNA and 5S-rRNA genes lie entirely within the transcribed sequence (Figure 11-47). Two such *internal* promoter elements, termed the *A box* and *B box*, are present in all tRNA genes. These highly conserved sequences not only function as promoters but also encode two invariant portions of eukaryotic tRNAs that are required for protein synthesis. In 5S-rRNA genes, a single internal control region, the *C box*, acts as a promoter.

Three general transcription factors are required for Pol III to initiate transcription of tRNA and 5S-rRNA genes *in vitro*. Two multimeric factors, TFIIB and TFIIC, participate in initiation at both tRNA and 5S-rRNA promoters; a third factor, TFIIA, is required for initiation at 5S-rRNA promoters. As with assembly of Pol I and Pol II initiation complexes, the Pol III general transcription factors bind to promoter DNA in a defined sequence.

The N-terminal half of one TFIIB subunit, called *BRF* (for TFIIB-related factor), is similar in sequence to TFIIB (a Pol II factor). This similarity suggests that BRF and TFIIB perform a similar function in initiation, namely, to direct the polymerase to the correct start site. Once TFIIB has bound to either a tRNA or 5S-rRNA gene, Pol III can bind and initiate transcription in the presence of ribonucleoside triphosphates. The BRF subunit of TFIIB interacts specifically with one of the polymerase subunits unique to Pol III, accounting for initiation by this specific nuclear RNA polymerase.

Another of the three subunits composing TFIIB is TBP, which we can now see is a component of a general transcription factor for all three eukaryotic nuclear RNA polymerases. The finding that TBP participates in transcription initiation by Pol I and Pol III was surprising, since the promoters recognized by these enzymes often do not contain TATA boxes. Nonetheless, recent studies indicate that the TBP subunit of TFIIB interacts with DNA similarly to the way it interacts with TATA boxes.



▲ **FIGURE 11-47 Transcription-control elements in genes transcribed by RNA polymerase III.** Both tRNA and 5S-rRNA genes contain internal promoter elements (yellow) located downstream from the start site and named A-, B-, and C-boxes, as indicated. Assembly of transcription initiation complexes on these genes begins with the binding of Pol III-specific general transcription factors TFIIA, TFIIB, and TFIIC to these control elements. Green arrows indicate strong, sequence-specific protein-DNA interactions. Blue arrows indicate interactions between general transcription factors. Purple arrows indicate interactions between general transcription factors and Pol III. [From L. Schramm and N. Hernandez, 2002, *Genes Dev.* 16:2593.]

Mitochondrial and Chloroplast DNAs Are Transcribed by Organelle-Specific RNA Polymerases

As discussed in Chapter 10, mitochondria and chloroplasts probably evolved from bacteria that were endocytosed into ancestral cells containing a eukaryotic nucleus. In modern-day eukaryotes, both organelles contain distinct circular DNAs that encode some of the proteins essential to their specific functions. The RNA polymerases that transcribe mitochondrial (mt) DNA and chloroplast DNA are similar to polymerases from bacteria and bacteriophages.

Mitochondrial RNA Polymerase The RNA polymerase that transcribes mtDNA is encoded in nuclear DNA. After synthesis of the enzyme in the cytosol, it is imported into the mitochondrial matrix by mechanisms described in Chapter 16. The mitochondrial RNA polymerases from *S. cerevisiae* and the frog *Xenopus laevis* both consist of a large subunit with ribonucleotide-polymerizing activity and a small subunit, or specificity factor, essential for initiating transcription at the start sites in mtDNA used in the cell. The large subunit of yeast mitochondrial RNA polymerase clearly is related to the monomeric RNA polymerases of bacteriophage T7 and similar bacteriophages. However, the mitochondrial enzyme is functionally distinct from the bacteriophage enzyme in its dependence on the small subunit for transcription from the proper start sites. This small subunit is related to the σ factors in bacterial RNA polymerases, which interact with promoter DNA and function as initiation factors. Thus mitochondrial RNA polymerase appears to be a hybrid of the simple bacteriophage RNA polymerases and the multisubunit bacterial RNA polymerases of intermediate complexity.

The promoter sequences recognized by mitochondrial RNA polymerases include the transcription start site. These promoter sequences, which are rich in A residues, have been characterized in the mtDNA from yeast, plants, and animals. The circular, human mitochondrial genome contains two related 15-bp promoter sequences, one for the transcription of each strand. Each strand is transcribed in its entirety; the long primary transcripts are then processed to yield mitochondrial mRNAs, rRNAs, and tRNAs. A small basic protein called *mtTF1*, which binds immediately upstream from the two mitochondrial promoters, greatly stimulates transcription. A homologous protein found in yeast mitochondria is required for maintenance of mtDNA and probably performs a similar function.

Chloroplast RNA Polymerase In contrast to mitochondrial RNA polymerase, the enzyme that transcribes chloroplast DNA is encoded in the chloroplast genome itself. This RNA polymerase has subunits with considerable homology to the *E. coli* RNA polymerase α , β , and β' subunits, but apparently lacks a subunit equivalent to the *E. coli* σ factor.

Some chloroplast promoters are quite reminiscent of the *E. coli* σ^{70} -promoter, with similar sequences in the -10 and -35 regions. Transcription from one chloroplast promoter, however, depends on sequences from about -20 to $+60$, quite different from most *E. coli* promoters. This promoter may be recognized by a second RNA polymerase, which most likely is encoded in the nuclear genome and imported into the organelle. Analysis of chloroplast transcription is still in its infancy, but at this point it is clear that at least one transcription system is highly homologous to transcription in *E. coli* and other bacteria from which chloroplasts evolved.

KEY CONCEPTS FOR SECTION 11.8

Other Eukaryotic Transcription Systems

- The process of transcription initiation by Pol I and Pol III is similar to that by Pol II but requires different general transcription factors, is directed by different promoter elements, and does not require ATP hydrolysis.
- Mitochondrial DNA is transcribed by a nuclear-encoded RNA polymerase composed of two subunits. One subunit is homologous to the monomeric RNA polymerase from bacteriophage T7; the other resembles bacterial σ factors.
- Chloroplast DNA is transcribed by a chloroplast-encoded RNA polymerase homologous to bacterial RNA polymerases, except that it lacks a σ factor.

PERSPECTIVES FOR THE FUTURE

A great deal has been learned in recent years about transcription control in eukaryotes. Genes encoding about 2000 activators and repressors can be recognized in the human genome. We now have a glimpse of how the astronomical number of possible combinations of these transcription factors can generate the complexity of gene control required to produce organisms as remarkable as those we see around us. But very much remains to be understood. While we now have some understanding of what processes turn a gene on and off, we have very little understanding of how the frequency of transcription is controlled in order to provide a cell with the appropriate amounts of its various proteins. In a red blood cell precursor, for example, the globin genes are transcribed at a far greater rate than the genes encoding the enzymes of intermediary metabolism. How are the vast differences in the frequency of transcription initiation at various genes achieved? What happens to the multiple interactions between activation domains, co-activator complexes, general transcription factors, and RNA polymerase II when the polymerase initiates transcription and transcribes away from the promoter region? Do these completely dissociate at promoters that are transcribed infrequently, so that the combination of multiple factors required for transcription must be

reassembled anew for each round of transcription? Do complexes of activators with their multiple interacting co-activators remain assembled at promoters from which reinitiation takes place at a high rate, so that the entire assembly does not have to be reconstructed each time a polymerase initiates?

Much remains to be learned about the structure of chromatin and how that structure influences transcription. What directs certain regions of chromatin to form heterochromatin where transcription is repressed? Precisely how is the structure of chromatin changed by activators and repressors and how does this promote or inhibit transcription? Once chromatin-remodeling complexes and histone acetylase complexes become associated with a promoter region, how do they remain associated? Do certain subunits of these complexes associate with modified histone tails so that additional complexes associate along the length of a chromatin fiber as additional histones are modified?

Single activation domains have been discovered to interact with several co-activator complexes. Are these interactions transient, so that the same activation domain can interact with several co-activators sequentially? Is a specific order of co-activator interaction required? How does the interaction of activation domains with mediator stimulate transcription? Do these interactions simply stimulate the assembly of a preinitiation complex, or do they also influence the rate at which RNA polymerase II initiates transcription from an assembled preinitiation complex?

Transcriptional activation is a highly cooperative process so that genes expressed in a specific type of cell are expressed only when the complete set of activators that control that gene are expressed and activated. As mentioned earlier, some of the transcription factors that control expression of the *TTR* gene in the liver are also expressed in intestinal and kidney cells. Yet the *TTR* gene is not expressed in these other tissues, since its transcription requires two additional transcription factors expressed only in the liver. What mechanisms account for this highly cooperative action of transcription factors that is critical to cell-type-specific gene expression?

A thorough understanding of normal development and of abnormal processes associated with disease will require answers to these and many other questions. As further understanding of the principles of transcription control are discovered, applications of the knowledge will likely be made. This understanding may allow fine control of the expression of therapeutic genes introduced by gene therapy vectors as they are developed. Detailed understanding of the molecular interactions that regulate transcription may provide new targets for the development of therapeutic drugs that inhibit or stimulate the expression of specific genes. A more complete understanding of the mechanisms of transcriptional control may allow improved engineering of crops with desirable characteristics. Certainly, further advances in the area of transcription control will help to satisfy our desire to understand how complex organisms such as ourselves develop and function.

KEY TERMS

activation domain 461	<i>MAT</i> locus (in yeast) 471
activators 449	mediator 471
antitermination factor 485	nuclear receptors 464
carboxyl-terminal domain (CTD) 452	promoter 449
chromatin-mediated repression 471	promoter-proximal elements 455
co-activator 467	repression domain 462
DNase I footprinting 458	repressors 449
enhancers 456	RNA polymerase II 450
enhancesome 468	silencer sequences 472
general transcription factors 469	TATA box 454
heat-shock genes 486	upstream activating sequences (UASs) 457
histone code 476	yeast two-hybrid system 480
histone deacetylation 474	zinc finger 463
leucine zipper 464	

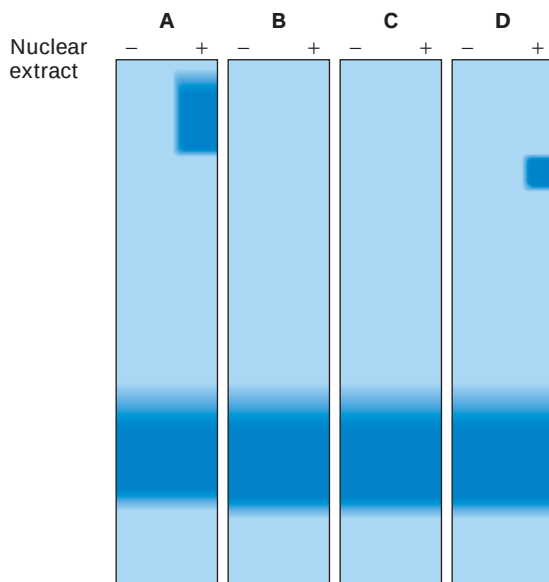
REVIEW THE CONCEPTS

1. What is the evidence that transcriptional initiation is the primary mechanism of gene control in complex organisms?
2. What types of genes are transcribed by RNA polymerases I, II, and III? Design an experiment to determine whether a specific gene is transcribed by RNA polymerase II.
3. The CTD of the largest subunit of RNA polymerase II can be phosphorylated and hyperphosphorylated at various serine and tyrosine residues. What are the conditions that lead to phosphorylation versus hyperphosphorylation?
4. What do TATA boxes, initiators, and CpG islands have in common? Which was the first of these to be identified? Why?
5. Describe the methods used to identify the location of DNA control elements in regulatory regions of genes.
6. What is the difference between a promoter-proximal element and a distal enhancer?
7. Describe the methods used to identify the location of DNA-binding proteins in the regulatory regions of genes.
8. Describe the structural features of transcriptional activator and repressor proteins.
9. What happens to transcription of the *EGR-1* gene in patients with Wilm's tumor? Why?
10. Using CREB and nuclear receptors as examples, compare and contrast the structural changes that take place when these transcription factors bind to their co-activators.

11. What structural change takes place on polymerase II promoters during preinitiation complex formation?
12. Expression of recombinant proteins in yeast is an important tool for biotechnology companies that produce new drugs for human use. In an attempt to get a new gene *X* expressed in yeast, a researcher has integrated gene *X* into the yeast genome near a telomere. Will this strategy result in good expression of gene *X*? Why or why not? Would the outcome of this experiment differ if the experiment had been performed in a yeast line containing mutations in the H3 or H4 histone tails?
13. You have isolated a new protein called STICKY. You can predict from comparisons with other known proteins that STICKY contains a bHLH domain and a Sin3-interacting domain. Predict the function of STICKY and rationale for the importance of these domains in STICKY function.
14. Describe at least one gene you would expect to be able to clone using the following genes as bait in a yeast two-hybrid experiment: alpha-globin; the catalytic subunit of protein kinase A; and the catalytic subunit of aspartate transcarbamylase.

ANALYZE THE DATA

An electrophoretic mobility shift assay (EMSA) was performed using a radiolabeled DNA fragment from the sequence upstream of gene *X*. This DNA probe was incubated with (+) or without (–) nuclear extract isolated from tissues A (bone); B (lung); C (brain); and D (skin). The DNA:protein complexes were then fractionated on nondenaturing polyacrylamide gels. The gels were exposed to autoradiographic film; the results are presented in the figure.



- a. Which tissues contain a binding activity that recognizes the sequence upstream of gene *X*? Is the transcription factor the same in each tissue?
- b. If the binding activity was purified, what test could be done to verify that this factor is in fact a transcription factor?
- c. What type of assay would be performed to determine the specific DNA sequence(s) to which the transcription factor binds?
- d. If gene *X* is transcribed in lung and brain tissue but not in bone and skin tissue, what type of transcription factor is the binding activity? Speculate as to the identity of other factors that might be complexed at the gene *X* promoter in bone and skin tissue.

REFERENCES

Overview of Eukaryotic Gene Control and RNA Polymerases

Cramer, P. 2002. Multisubunit RNA polymerases. *Curr. Opin. Struc. Biol.* **12**:89–97.

Regulatory Sequences in Protein-Coding Genes

Blackwood, E. M., and J. T. Kadonaga. 1998. Going the distance: a current view of enhancer action. *Science* **281**:60–63.

Butler, J. E., and J. T. Kadonaga. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16**:2583–2592.

http://www.epd.isb-sib.ch/promoter_elements

Activators and Repressors of Transcription

Brivanlou, A. H., and J. E. Darnell, Jr. 2002. Signal transduction and the control of gene expression. *Science* **295**:813–818.

Luscombe, N. M., et al. 2000. An overview of the structures of protein-DNA complexes. *Genome Biol.* **1**:1–37.

Riechmann, J. L., et al. 2000. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**:2105–2110.

Tupler, R., G. Perini, and M. R. Green. 2001. Expressing the human genome. *Nature* **409**:832–833.

Transcription Initiation by RNA Polymerase II

Woychik, N. A., and M. Hampsey. 2002. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**:453–463.

Molecular Mechanisms of Transcription Activation and Repression

Berger, S. L. 2002. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**:142–148.

Boube, M., et al. 2002. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* **110**:143–151.

Courey, A. J., and S. Jia. 2001. Transcriptional repression: the long and the short of it. *Genes Dev.* **15**:2786–2796.

Horn, P. J., and C. L. Peterson. 2002. Chromatin Higher Order Folding—Wrapping up Transcription. *Science* **297**:1824–1827.

Richards, E. J., and S. C. Elgin. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**:489–500.

Control of Transcription-Factor Activity

Hermanson, O., C. K. Glass, and M. G. Rosenfeld. 2002. Nuclear receptor coregulators: multiple modes of modification. *Trends Endocrinol. Metab.* **13**:55–60.

Lin, R. J., et al. 1998. The transcriptional basis of steroid physiology. *Cold Spring Harbor Symp. Quant. Biol.* **63**:577–585.

McKenna, N. J., and B. W. O'Malley. 2002. Minireview: nuclear receptor coactivators—an update. *Endocrinol.* **143**:2461–2465.

Yamamoto, K. R. 1995. Multilayered control of intracellular receptor function. *Harvey Lect.* **91**:1–19.

Regulated Elongation and Termination of Transcription

Garber, M. E., and K. A. Jones. 1999. HIV-1 Tat: coping with negative elongation factors. *Curr. Opin. Immunol.* **11**:460–465.

Kim, D. K., et al. 2001. The regulation of elongation by eukaryotic RNA polymerase II: a recent view. *Mol. Cells* **11**:267–274.

Lis, J. 1998. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harbor Symp. Quant. Biol.* **63**:347–356.

Price, D. H. 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell Biol.* **20**:2629–2634.

Other Eukaryotic Transcription Systems

Paule, M. R., and R. J. White. 2000. Survey and summary: transcription by RNA polymerases I and III. *Nucl. Acids Res.* **28**:1283–1298.

Schramm, L., and N. Hernandez. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**:2593–2620.