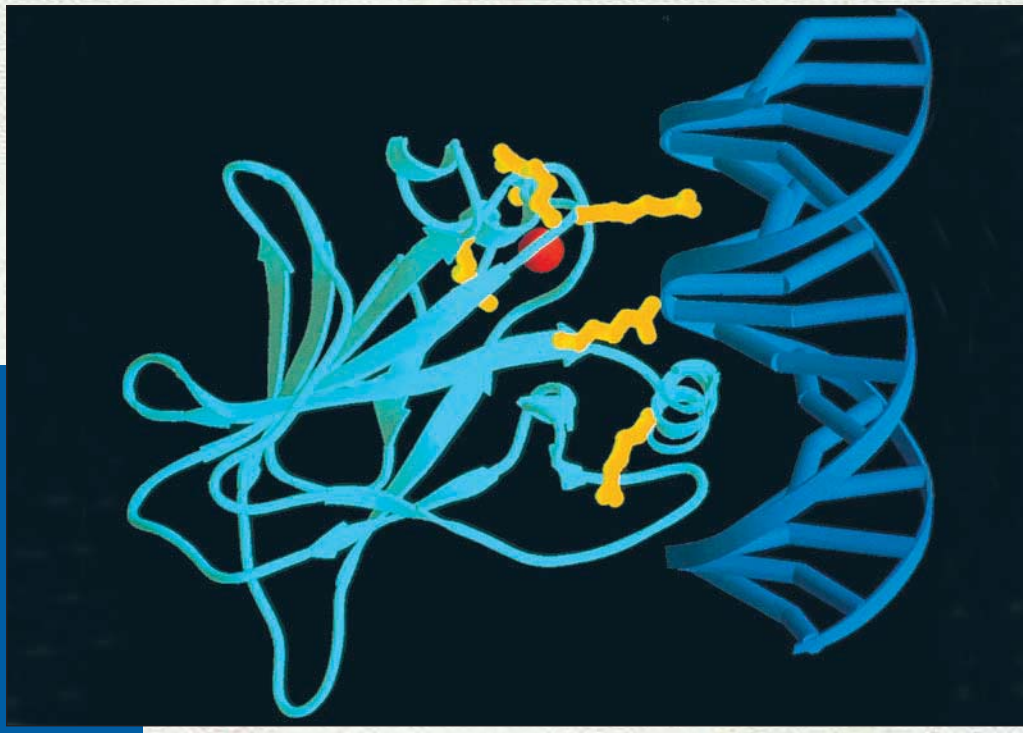
CHAPTER 12

Transcription Activators in Eukaryotes



Computer model of the transcription factor p53 interacting with its target DNA site. *Courtesy Nicola P. Pavletich, Sloan-Kettering Cancer Center, Science (15 July 1994) cover. Copyright © AAAS.*

In Chapters 10 and 11 we learned about the basic machinery involved in eukaryotic transcription: the three RNA polymerases, their promoters, and the general transcription factors that bring RNA polymerase and promoter together. However, it is clear that this is not the whole story. The general transcription factors by themselves dictate the starting point and direction of transcription, but they are capable of sponsoring only a very low level of transcription (basal level transcription). But transcription of active genes in cells rises above (frequently far above) the basal level. To provide the needed extra boost in transcription, eukaryotic cells have additional, gene-specific transcription factors **(activators)** that bind to DNA elements called enhancers (Chapter 10). The transcription activation provided by these activators also permits cells to control the expression of their genes.

In addition, eukaryotic DNA is complexed with protein in a structure called chromatin. Some chromatin, called heterochromatin, is highly condensed and inaccessible to RNA polymerases, so it cannot be transcribed. Other chromatin (euchromatin) still contains protein, but it is relatively extended. Much of this euchromatin, even though it is relatively open, contains genes that are not transcribed in a given cell because the appropriate activators are not available to turn them on. Instead, other proteins may hide the promoters from RNA polymerase and general transcription factors to ensure that they remain turned off. In this chapter, we will examine the activators that control eukaryotic genes. Then, in Chapter 13, we will look at the crucial relationship among activators, chromatin structure, and gene activity.

12.1 Categories of Activators

Activators can either stimulate or inhibit transcription by RNA polymerase II, and they have structures composed of at least two functional domains: a **DNA-binding domain** and a **transcription-activating domain.** Many also have a **dimerization domain** that allows the activators to bind to each other, forming homodimers (two identical monomers bound together), heterodimers (two different monomers bound together), or even higher multimers such as tetramers. Some even have binding sites for effector molecules like steroid hormones. Let us consider some examples of these three kinds of structural–functional domains, bearing in mind an important principle we discussed in Chapters 6 and 9: A protein does not have just one shape. Rather, it is a dynamic molecule that assumes many possible conformations. Some of these may be especially advantageous for binding to other molecules, such as a specific DNA sequence, and these conformations would be stabilized by binding to such DNA sequences. Thus, when we refer to the shape of a DNA-binding protein, or a domain within such a protein, we mean one of many possible shapes, which happens to fit particularly well with the DNA in question.

DNA-Binding Domains

A protein **domain** is an independently folded region of a protein. Each DNA-binding domain has a **DNA-binding motif,** which is the part of the domain that has a characteristic shape specialized for specific DNA binding. Most DNA-binding motifs fall into the following classes:

1. *Zinc-containing modules.* At least three kinds of zinc-containing modules act as DNA-binding motifs. These all use one or more zinc ions to create the proper shape so an α-helix within the motif can fit into the DNA major groove and make specific contacts there. These zinc-containing modules include:
   a. **Zinc fingers,** such as those found in TFIIIA and Sp1, two transcription factors we have already encountered.
   b. Zinc modules found in the glucocorticoid receptor and other members of this group of nuclear receptors.
   c. Modules containing two zinc ions and six cysteines, found in the yeast activator GAL4 and its relatives.
2. **Homeodomains (HDs).** These contain about 60 amino acids and resemble in structure and function the helix-turn-helix DNA-binding domains of prokaryotic proteins such as the λ phage repressor. HDs, found in a variety of activators, were originally identified in activators called homeobox proteins that regulate development in the fruit fly *Drosophila*.
3. **bZIP and bHLH motifs.** The CCAAT/enhancer-binding protein (C/EBP), the MyoD protein, and many other eukaryotic transcription factors have a highly basic DNA-binding motif linked to one or both of the protein dimerization motifs known as leucine zippers and helix-loop-helix (HLH) motifs. (By the way C/EBP is different from the CCAAT-binding transcription factor [CTF, Chapter 10]).

This list is certainly not exhaustive. In fact, several transcription factors have now been identified that do not fall into any of these categories.

Transcription-Activating Domains

Most activators have one of these domains, but some have more than one. So far, most of these domains fall into three classes, as follows:

1. *Acidic domains.* The yeast activator GAL4 typifies this group. It has a 49-amino-acid domain with 11 acidic amino acids.
2. *Glutamine-rich domains.* The activator Sp1 has two such domains, which are about 25% glutamine. One of these has 39 glutamines in a span of 143 amino acids. In addition, Sp1 has two other activating domains that do not fit into any of these three main categories.
3. *Proline-rich domains.* The activator CTF, for instance, has a domain of 84 amino acids, 19 of which are prolines.

Our descriptions of the transcription-activating domains are necessarily nebulous, because the domains themselves are rather ill-defined. The acidic domain, for example, has seemed to require nothing more than a preponderance of acidic residues to make it function, which led to the name "acid blob" to describe this presumably unstructured

domain. On the other hand, Stephen Johnston and his colleagues have shown that the acidic activation domain of GAL4 tends to form a defined structure—a β-sheet—in slightly acidic solution. It is possible that the β-sheet also forms under the slightly basic conditions in vivo, but this is not yet clear. These workers also removed all six of the acidic amino acids in the GAL4 acidic domain and showed that it still retained 35% of its normal ability to activate transcription. Thus, not only is the structure of the acidic activating domain unclear, the importance of its acidic nature is even in doubt.

With such persistent uncertainty, it has been difficult to draw conclusions about how the structure and function of transcription-activating domains are related. On the other hand, some evidence suggests that the glutamine-rich activation domain of Spl operates by interacting with glutamine-rich domains of other transcription factors.

SUMMARY Eukaryotic activators are composed of at least two domains: a DNA-binding domain and a transcription-activating domain. DNA-binding domains contain motifs such as zinc modules, homeodomains, and bZIP or bHLH motifs. Transcription-activating domains can be acidic, glutamine-rich, or proline-rich.

12.2 Structures of the DNA-Binding Motifs of Activators

By contrast to the transcription-activating domains, most DNA-binding domains have well-defined structures, and x-ray crystallography studies have shown how these structures interact with their DNA targets. Furthermore, these same structural studies have frequently elucidated the dimerization domains responsible for interaction between protein monomers to form a functional dimer, or in some cases, a tetramer. This is crucial, because most classes of DNA-binding proteins are incapable of binding to DNA in monomer form; they must form at least dimers to function. Let us explore the structures of several classes of DNA-binding motifs and see how they mediate interaction with DNA. In the process we will discover the ways some of these proteins can dimerize.

Zinc Fingers

In 1985, Aaron Klug noticed a periodicity in the structure of the general transcription factor TFIIIA. This protein has nine repeats of a 30-residue element. Each element has two closely spaced cysteines followed 12 amino acids later by two closely spaced histidines. Furthermore, the protein is rich in zinc—enough for one zinc ion per repeat. This led Klug to predict that each zinc ion is complexed by the two cysteines and two histidines in each repeat unit to form a finger-shaped domain.

Finger Structure Michael Pique and Peter Wright used nuclear magnetic resonance spectroscopy to determine the structure in solution of one of the zinc fingers of the *Xenopus laevis* protein Xfin, an activator of certain class II promoters. Note that this structure, depicted in Figure 12.1, really is not very finger-shaped, unless it is a rather wide, stubby finger. It is also worth noting that this finger shape by itself does not confer any binding specificity, since there are many different finger proteins, all with the same shape fingers but each binding to its own unique DNA target sequence. Thus, it is the precise amino acid sequences of the fingers, or of neighboring parts of the protein, that determine the DNA sequence to which the protein can bind. In the Xfin finger, an α-helix (on the left in Figure 12.1) contains several basic amino acids—all on the side that seems to contact the DNA. These and other amino acids in the helix presumably determine the binding specificity of the protein.

Carl Pabo and his colleagues used x-ray crystallography to obtain the structure of the complex between DNA,

Figure 12.1 Three-dimensional structure of one of the zinc fingers of the *Xenopus* protein Xfin. The zinc is represented by the turquoise sphere at top center. The sulfurs of the two cysteines are represented by yellow-green spheres. The two histidines are represented by the blue-green structures at upper left. The backbone of the finger is represented by the purple tube. (*Source:* Pique, Michael and Peter E. Wright, Dept. of Molecular Biology, Scripps Clinic Research Institute, La Jolla, CA. (cover photo, *Science* 245 (11 Aug 1989).)

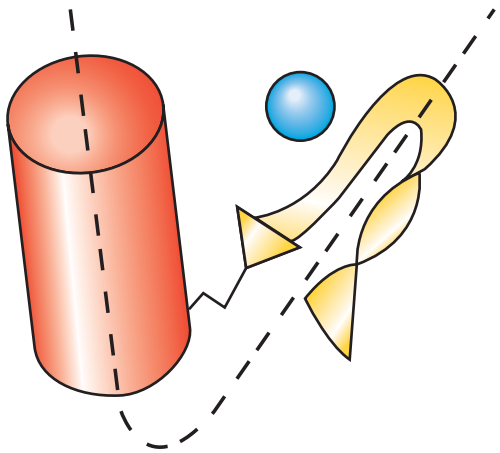


Figure 12.2 Schematic diagram of zinc finger 1 of the Zif268 protein. The right-hand side of the finger is an antiparallel β-sheet (yellow), and the left-hand side is an α-helix (red). Two cysteines in the β-sheet and two histidines in the α-helix coordinate the zinc ion in the middle (blue). The dashed line traces the outline of the "finger" shape. (*Source:* Adapted from Pavletich, N.P. and C.O. Pabo, Zinc finger–DNA recognition: Crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* 252:812, 1991.)

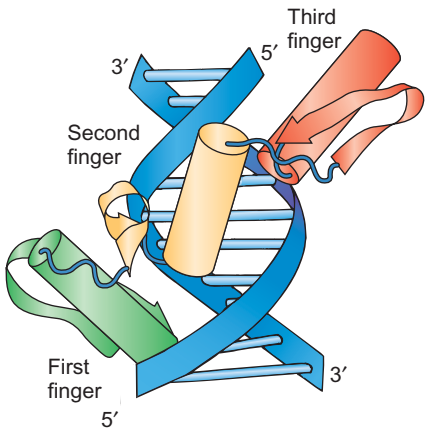


Figure 12.3 Arrangement of the three zinc fingers of Zif268 in a curved shape to fit into the major groove of DNA. As usual, the cylinders and ribbons stand for α-helices and β-sheets, respectively. (*Source:* Adapted from Pavletich, N.P. and C.O. Pabo, Zinc finger–DNA recognition: Crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* 252:811, 1991.)

and a member of the TFIIIA class of zinc finger proteins—the mouse protein Zif268. This is a so-called *immediate early protein,* which means that it is one of the first genes to be activated when resting cells are stimulated to divide. The Zif268 protein has three adjacent zinc fingers that fit into the major groove of the DNA double helix. We will see the arrangement of these three fingers a little later in the chapter. For now, let us consider the three-dimensional structure of the fingers themselves. Figure 12.2 presents the structure of finger 1 as an example. The finger shape in this presentation is perhaps not obvious. Still, on close inspection we can see the finger contour, which is indicated by the dashed line. As in the Xfin zinc finger, the left side of each Zif268 finger is an α-helix. This is connected by a short loop at the bottom to the right side of the finger, a small antiparallel β-sheet. Do not confuse this β-sheet itself with the finger; it is only one half of it. The zinc ion (blue sphere) is in the middle, coordinated by two histidines in the α-helix and by two cysteines in the β-sheet. All three fingers have almost exactly the same shape.

Interaction with DNA How do the fingers interact with their DNA targets? Figure 12.3 shows all three Zif268 fingers lining up in the major groove of the DNA. In fact, the three fingers are arranged in a curve, or C-shape, which matches the curve of the DNA double helix. All the fingers approach the DNA from essentially the same angle, so the geometry of protein–DNA contact is very similar in each case. Binding between each finger and its DNA-binding site relies on direct amino acid–base interactions, between amino acids in the α-helix and bases in the major groove of the DNA. For more detailed descriptions of amino acid–base interactions, see Chapter 9.

Comparison with Other DNA-Binding Proteins One unifying theme emerging from studies of many, but not all, DNA-binding proteins is the utility of the α-helix in contacting the DNA major groove. We saw many examples of this with the prokaryotic helix-turn-helix domains (Chapter 9), and we will see several other eukaryotic examples. What about the β-sheet in Zif268? It seems to serve the same function as the first α-helix in a helix-turn-helix protein, namely to bind to the DNA backbone and help position the recognition helix for optimal interaction with the DNA major groove.

Zif268 also shows some differences from the helix-turn-helix proteins. Whereas the latter proteins have a single DNA-binding domain per monomer, the finger protein DNA-binding domains have a modular construction, with several fingers making contact with the DNA. This arrangement means that these proteins, in contrast to most DNA-binding proteins, do not need to form dimers or tetramers to bind to DNA. They already have multiple binding domains built in. Also, most of the protein–DNA contacts are with one DNA strand, rather than both, as in the case of the helix-turn-helix proteins. At least with this particular finger protein, most of the contacts are with bases, rather than the DNA backbone.

In 1991, Nikola Pavletich and Carl Pabo solved the structure of a cocrystal between DNA and a five-zinc-finger human protein called GLI. This provided an interesting contrast with the three-finger Zif268 protein. Again, the major groove is the site of finger–DNA contacts, but in this case one finger (finger 1) does not contact the DNA. Also, the overall geometries of the two finger–DNA complexes are similar, with the fingers wrapping around the DNA

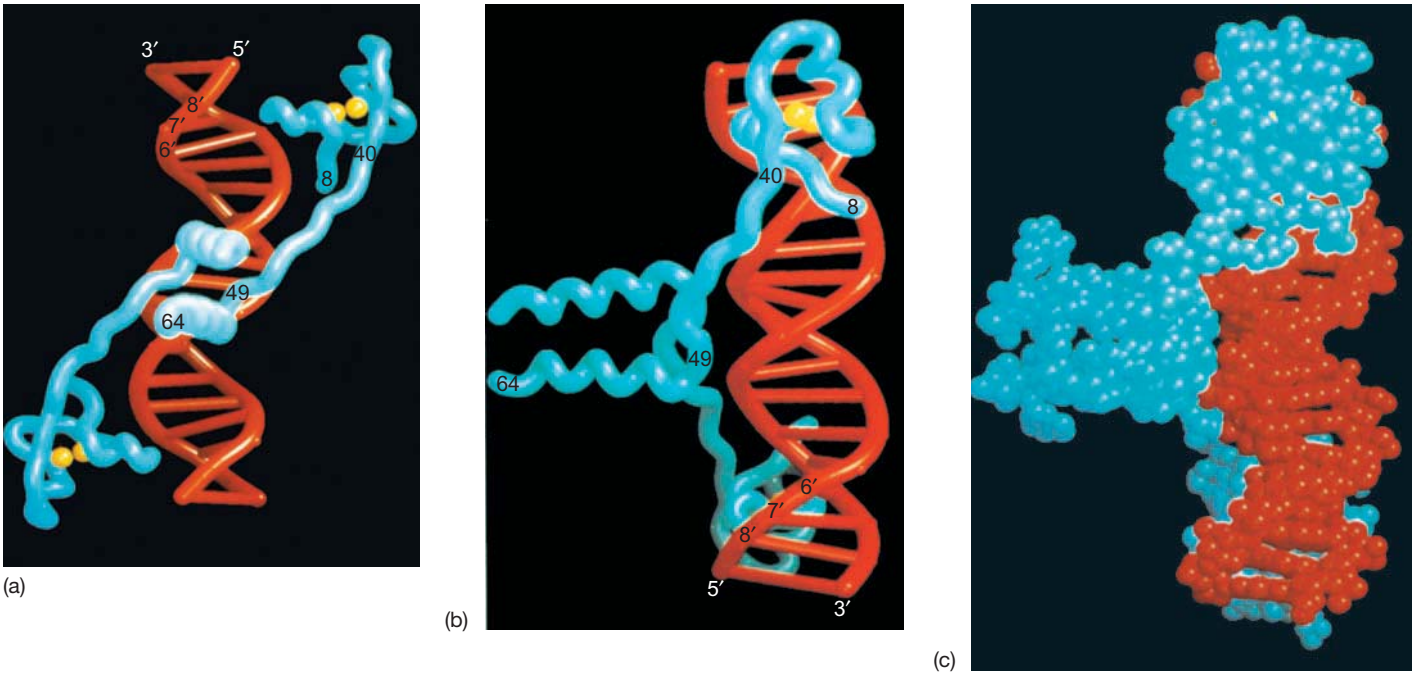


Figure 12.4 Three views of the GAL4–DNA complex. (a) The complex viewed approximately along its two-fold axis of symmetry. The DNA is in red, the protein is in blue, and the zinc ions are represented by yellow spheres. Amino acid residue numbers at the beginnings and ends of the three domains are given on the top monomer: The DNA recognition module extends from residue 8 to 40. The linker, from residue 41 to 49, and the dimerization domain, from residue 50 to 64. **(b)** The complex viewed approximately perpendicular to the view in panel (a). The dimerization elements appear roughly parallel to one another at left center. **(c)** Space-filling model of the complex in the same orientation as in panel (b). Notice that the recognition modules on the two GAL4 monomers make contact with opposite faces of the DNA. Notice also the neat fit between the coiled coil of the dimerization domain and the minor groove of the DNA helix. (*Source:* Marmorstein, R., M. Carey, M. Ptashne, and S.C. Harrison, DNA recognition by GAL4: Structure of a protein–DNA complex. *Nature* 356 (2 April 1992) p. 411, f. 3. Copyright © Macmillan Magazines Ltd.)

major groove, but no simple "code" of recognition between certain bases and amino acids exists.

SUMMARY Zinc fingers are composed of an antiparallel β-sheet, followed by an α-helix. The β-sheet contains two cysteines, and the α-helix two histidines, that are coordinated to a zinc ion. This coordination of amino acids to the metal helps form the finger-shaped structure. The specific recognition between the finger and its DNA target occurs in the major groove.

The GAL4 Protein

The **GAL4** protein is a yeast activator that controls a set of genes responsible for metabolism of galactose. Each of these GAL4-responsive genes contains a GAL4 target site (enhancer) upstream of the transcription start site. These target sites are called **upstream activating sequences,** or **UAS_Gs.** GAL4 binds to a UAS_G as a dimer. Its DNA-binding motif is located in the first 40 amino acids of the protein, and its dimerization motif is found in residues 50–94. The DNA-binding motif is similar to the zinc finger in that it contains zinc and cysteine residues, but its structure must be different: Each motif has six cysteines and no histidines, and the ratio of zinc ions to cysteines is 1:3.

Mark Ptashne and Stephen Harrison and their colleagues performed x-ray crystallography on cocrystals of the first 65 amino acids of GAL4 and a synthetic 17-bp piece of DNA. This revealed several important features of the protein–DNA complex, including the shape of the DNA-binding motif and how it interacts with its DNA target, and part of the dimerization motif in residues 50–64.

The DNA-Binding Motif Figure 12.4 depicts the structure of the GAL4 peptide dimer–DNA complex. One end of each monomer contains a DNA-binding motif containing six cysteines that complex two zinc ions (yellow spheres), forming a *bimetal thiolate cluster.* Each of these motifs also features a short α-helix that protrudes into the major groove of the DNA double helix, where its amino acid side chains can make specific interactions with the DNA bases and backbone. The other end of each monomer is an α-helix that serves a dimerization function that we will discuss later in this chapter.

The Dimerization Motif The GAL4 monomers also take advantage of α-helices in their dimerization, forming a

parallel **coiled coil** as illustrated at left in Figure 12.4b and c. This figure also shows that the dimerizing α-helices point directly at the minor groove of the DNA. Finally, note in Figure 12.4 that the DNA recognition module and the dimerization module in each monomer are joined by an extended linker domain. We will see other examples of coiled coil dimerization motifs when we discuss bZIP and bHLH motifs later in this chapter.

SUMMARY The GAL4 protein is a member of the zinc-containing family of DNA-binding proteins, but it does not have zinc fingers. Instead, each GAL4 monomer contains a DNA-binding motif with six cysteines that coordinate two zinc ions in a bimetal thiolate cluster. The recognition module contains a short α-helix that protrudes into the DNA major groove and makes specific interactions there. The GAL4 monomer also contains an α-helical dimerization motif that forms a parallel coiled coil as it interacts with the α-helix on the other GAL4 monomer.

The Nuclear Receptors

A third class of zinc module is found in the **nuclear receptors.** These proteins interact with a variety of endocrine-signaling molecules (steroids and other hormones) that diffuse through the cell membrane. They form hormone-receptor complexes that function as activators by binding to enhancers, or **hormone response elements,** and stimulating transcription of their associated genes. Thus, these activators differ from the others we have studied in that they must bind to an effector (a hormone) in order to function as activators. This implies that they must have an extra important domain—a hormone-binding domain—and indeed they do.

Some of the hormones that work this way are the sex hormones (androgens and estrogens); progesterone, the hormone of pregnancy (and principal ingredient of common birth control pills); the glucocorticoids, such as cortisol; vitamin D, which regulates calcium metabolism; and thyroid hormone and retinoic acid, which regulate gene expression during development. Each hormone binds to its specific receptor, and together they activate their own set of genes.

The nuclear receptors have traditionally been divided into three classes. The **type I receptors** include the steroid hormone receptors, typified by the **glucocorticoid receptor.** In the absence of their hormone ligands, these receptors reside in the cytoplasm, coupled with another protein. When a type I receptor binds to its hormone ligand, it releases its protein partner and migrates to the nucleus, where it binds as a homodimer to its hormone response element. For example, the glucocorticoid receptor exists in the cytoplasm complexed with a partner known as heat shock protein 90 (Hsp90). When the receptor binds to its glucocorticoid ligand (Figure 12.5), it changes conformation, dissociates from Hsp90, and moves into the nucleus

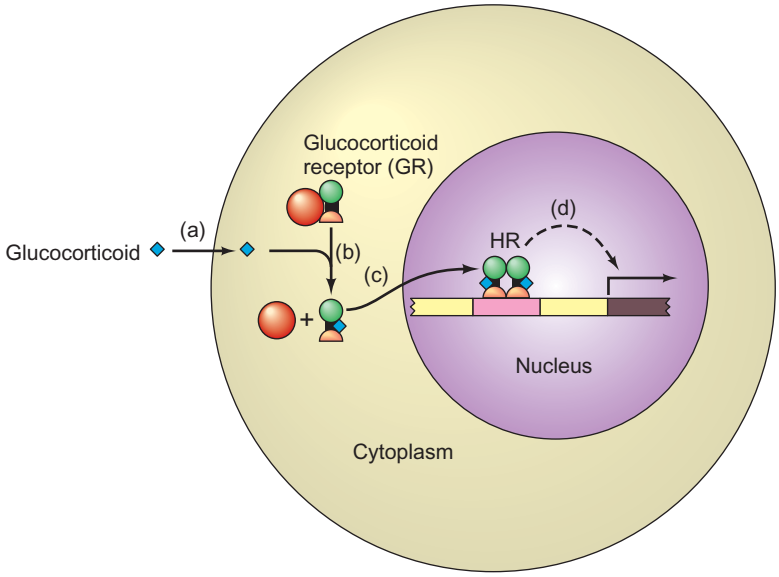


Figure 12.5 Glucocorticoid action. The glucocorticoid receptor (GR) exists in an inactive form in the cytoplasm complexed with heat shock protein 90 (Hsp90). **(a)** The glucocorticoid (blue diamond) diffuses across the cell membrane and enters the cytoplasm. **(b)** The glucocorticoid binds to its receptor (GR, red and green), which changes conformation and dissociates from Hsp90 (orange). **(c)** The hormone–receptor complex (HR) enters the nucleus, dimerizes with another HR, and binds to a hormone-response element, or enhancer (pink), upstream of a hormone-activated gene (brown). **(d)** Binding of the HR dimer to the enhancer activates (dashed arrow) the associated gene, so transcription occurs (bent arrow).

to activate genes controlled by enhancers called **glucocorticoid response elements (GREs).**

Sigler and colleagues performed x-ray crystallography on cocrystals of the glucocorticoid receptor and an oligonucleotide containing two target half-sites.

The crystal structure revealed several aspects of the protein–DNA interaction: (1) The binding domain dimerizes, with each monomer making specific contacts with one target half-site. (2) Each binding motif is a zinc module that contains two zinc ions, rather than the one found in a classical zinc finger. (3) Each zinc ion is complexed to four cysteines to form a finger-like shape. (4) The amino-terminal finger in each binding domain engages in most of the interactions with the DNA target. Most of these interactions involve an α-helix. The crystal structure revealed several aspects of the protein-DNA interaction: Figure 12.6 illustrates the specific amino-acid–base associations between this recognition helix and the DNA target site. Some amino acids outside this helix also make contact with the DNA through its backbone phosphates.

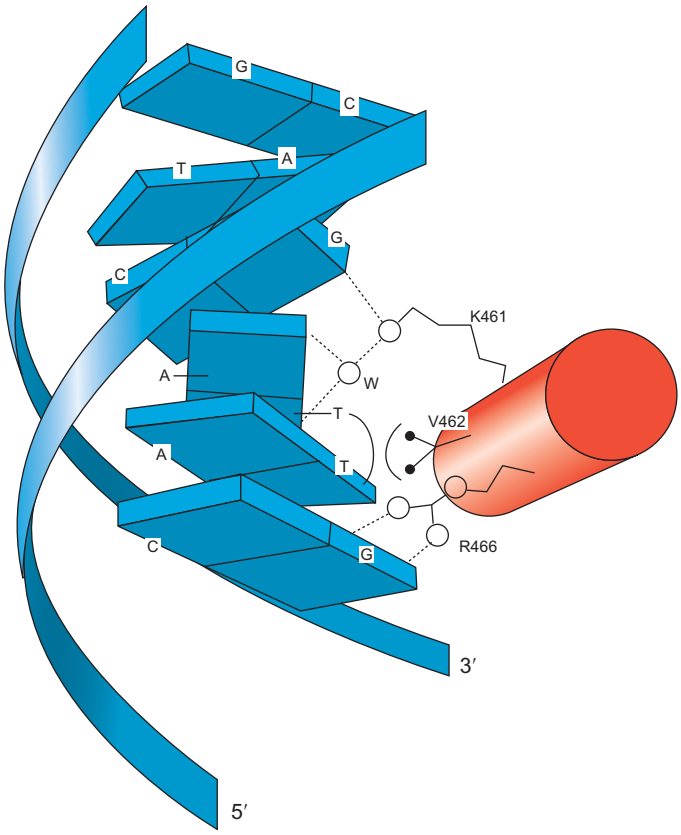


Figure 12.6 Association between the glucocorticoid receptor DNA-binding domain's recognition helix and its DNA target. The specific amino-acid–base interactions are shown. A water molecule (W) mediates some of the H-bonding between lysine 461 and the DNA. (*Source:* Adapted from Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler, Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352 (8 Aug 1991) p. 500, f. 4a. Copyright © Macmillan Magazines Ltd.)

The **type II receptors,** exemplified by the **thyroid hormone receptor,** stay in the nucleus, where they form dimers with another protein called **retinoic acid receptor X (RXR),** whose ligand is 9-*cis* retinoic acid. These receptors bind to their target sites in both the presence and absence of their ligands. As we will see in Chapter 13, binding of these type II receptors in the absence of ligand can repress transcription, whereas binding of the receptors along with their ligands can stimulate transcription. Thus, the same protein can act as either an activator or a repressor, depending on environmental conditions.

The **type III receptors** are not as well understood. They are also known as "orphan receptors" because their ligands have not been identified. Perhaps further study will show that some or all of these type III receptors really belong with the type I or type II receptors.

Finally, note that all three classes of zinc-containing DNA-binding modules use a common motif—an α-helix—for most of the interactions with their DNA targets.

SUMMARY Type I nuclear receptors reside in the cytoplasm, bound to another protein. When these receptors bind to their hormone ligands, they release their cytoplasmic protein partners and move to the nucleus where they bind to enhancers, and thereby act as activators. The glucocorticoid receptor is representative of this group. It has a DNA-binding domain with two zinc-containing modules. One module contains most of the DNA-binding residues (in a recognition α-helix), and the other module provides the surface for protein–protein interaction to form a dimer. Type II nuclear receptors, e.g., thyroid hormone receptor, stay in the nucleus, bound to their target DNA sites. In the absence of their ligands they repress gene activity, but when they bind their ligands they activate transcription. Type III receptors are "orphan" receptors whose ligands have not been identified.

Homeodomains

Homeodomains are DNA-binding domains found in a large family of activators. Their name comes from the gene regions, called **homeoboxes,** in which they are encoded. Homeoboxes were first discovered in regulatory genes of the fruit fly *Drosophila,* called homeotic genes. Mutations in these genes cause strange transformations of body parts in the fruit fly. For example, a mutation called *Antennapedia* causes legs to grow where antennae would normally be (Figure 12.7).

Homeodomain proteins are members of the helix-turn-helix family of DNA-binding proteins (Chapter 9). Each homeodomain contains three α-helices; the second and third of these form the helix-turn-helix motif, with the third

Figure 12.7 The *Antennapedia* phenotype. Legs appear on the head where antennae would normally be. (*Source:* Courtesy Walter J. Gehring, University of Basel, Switzerland.)

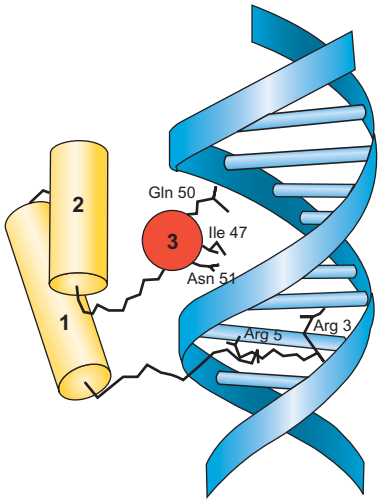


Figure 12.8 Representation of the homeodomain–DNA complex. Schematic model with the three helices numbered on the left, and a ribbon diagram of the DNA target on the right. The recognition helix (labeled 3, red) is shown on end, resting in the major groove of the DNA. The N-terminal arm is also shown, inserted into the DNA minor groove. Key amino acid side chains are shown interacting with DNA. (*Source:* Adapted from Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo, Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: A framework for understanding homeodomain–DNA interactions. *Cell* 63 (2 November, 1990) p. 582. f. 5b.)

serving as the recognition helix. But most homeodomains have another element, not found in helix-turn-helix motifs: The N-terminus of the protein forms an arm that inserts into the the minor groove of the DNA. Figure 12.8 shows the interaction between a typical homeodomain, from the *Drosophila* homeotic gene *engrailed,* and its DNA target. This view of the protein–DNA complex comes from Thomas Kornberg's and Carl Pabo's x-ray diffraction analysis of cocrystals of the engrailed homeodomain and an oligonucleotide containing the engrailed binding site. Most homeodomain proteins have weak DNA-binding specificity on their own. As a result, they rely on other proteins to help them bind specifically and efficiently to their DNA targets.

SUMMARY The homeodomains in eukaryotic activators contain a DNA-binding motif that functions in much the same way as helix-turn-helix motifs in which a recognition helix fits into the DNA major groove and makes specific contacts there. In addition, the N-terminal arm nestles in the adjacent minor groove.

The bZIP and bHLH Domains

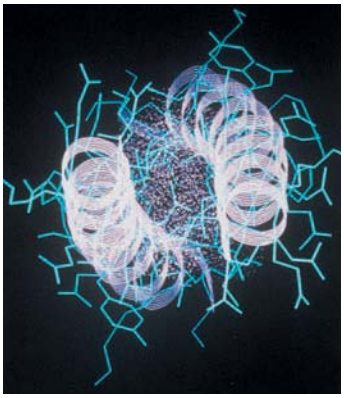
As with several of the other DNA-binding domains we have studied, the **bZIP** and **bHLH** domains combine two functions: DNA binding and dimerization. The *ZIP* and *HLH* parts of the names refer to the **leucine zipper** and **helix-loop-helix parts,** respectively, of the domains, which are the dimerization motifs. The *b* in the names refers to a basic region in each domain that forms the majority of the DNA-binding motif.

Let us consider the structures of these combined dimerization/DNA-binding domains, beginning with the bZIP domain. This domain actually consists of two polypeptides, each of which contains half of the zipper: an α-helix with leucine (or other hydrophobic amino acid) residues spaced seven amino acids apart, so they are all on one face of the helix. The spacing of the hydrophobic amino acids on one monomer puts them in position to interact with a similar string of amino acids on the other protein monomer. In this way, the two helices act like the two halves of a zipper.

To get a better idea of the structure of the zipper, Peter Kim and Tom Alber and their colleagues crystallized a synthetic peptide corresponding to the bZIP domain of GCN4, a yeast activator that regulates amino acid metabolism. The x-ray diffraction pattern shows that the dimerized bZIP domain assumes a parallel coiled coil structure (Figure 12.9). The α-helices are parallel in that their amino to carboxyl orientations are the same (left to right in panel b). Figure 12.9a, in which the coiled coil extends directly out at the reader, gives a good feel for the extent of supercoiling in the coiled coil. Notice the similarity between this and the coiled coil dimerization motif in GAL4 (see Figure 12.4).

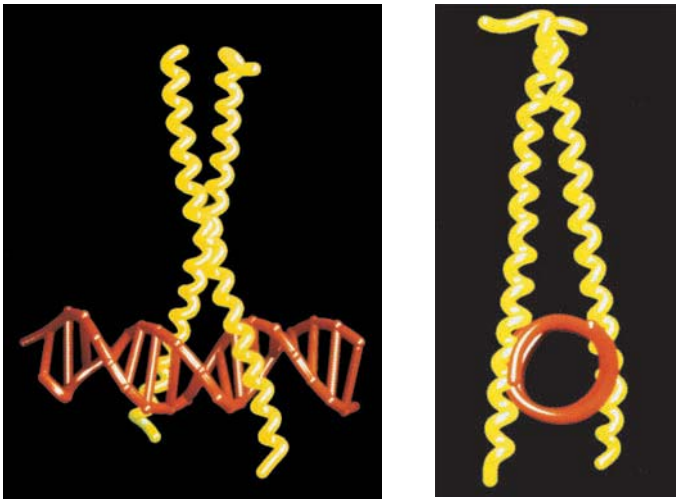
This crystallographic study, which focused on the zipper in the absence of DNA, did not shed light on the mechanism of DNA binding. However, Kevin Struhl and Stephen Harrison and their colleagues performed x-ray crystallography on the bZIP domain of GCN4, bound to its DNA target. Figure 12.10 shows that the leucine zipper not only brings the two monomers together, it also places the two basic parts of the domain in position to grasp the DNA like a pair of forceps, or fireplace tongs, with the basic motifs fitting into the DNA major groove.

(a)



(b)

Figure 12.9 Structure of a leucine zipper. (a) Kim and Alber and colleagues crystallized a 33-amino-acid peptide containing the leucine zipper motif of the transcription factor GCN4. X-ray crystallography on this peptide yielded this view along the axis of the zipper with the coiled coil pointed out of the plane of the paper. **(b)** A side view of the coiled coil with the two α-helices colored red and blue. Notice that the amino ends of both peptides are on the left. Thus, this is a parallel coiled coil. (*Source:* (*a*) O'Shea, E.K., J.D. Klemm, P.S. Kim, and T. Alber, X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254 (25 Oct 1991) p. 541, f. 3. Copyright © AAAS.)

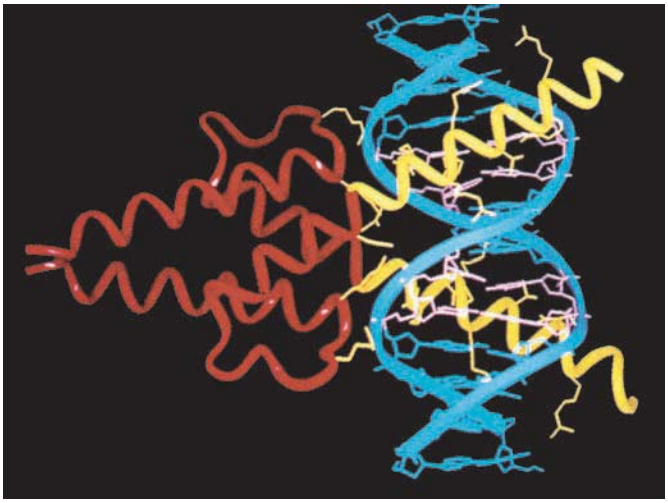


(a) (b)

Figure 12.10 Crystal structure of the bZIP motif of GCN4 bound to its DNA target. The DNA (red) contains a target for the bZIP motif (yellow). Notice the coiled coil nature of the interaction between the protein monomers, and the tong-like appearance of the protein grasping the DNA. **(a)** Side view of DNA. **(b)** End view of DNA. (*Source:* Ellenberger, T.E., C.J. Brandl, K. Struhl, and S.C. Harrison, The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: Crystal structure of the protein–DNA complex. *Cell* 71 (24 Dec 1992) p. 1227, f. 3a–b. Reprinted by permission of Elsevier Science.)

Harold Weintraub and Carl Pabo and colleagues solved the crystal structure of the bHLH domain of the activator **MyoD** bound to its DNA target. The structure (Figure 12.11) is remarkably similar to that of the bZIP domain–DNA complex we just considered. The helix-loop-helix part is the dimerization motif, but the long helix (helix 1) in each helix-loop-helix domain contains the basic region of the domain, which grips the DNA target via its major groove, just as the bZIP domain does.

Some proteins, such as the oncogene products Myc and Max, have **bHLH-ZIP domains** with both HLH and ZIP motifs adjacent to a basic motif. The bHLH-ZIP domains interact with DNA in a manner very similar to that employed by the bHLH domains. The main difference between bHLH and bHLH-ZIP domains is that the latter



(a)

(b)

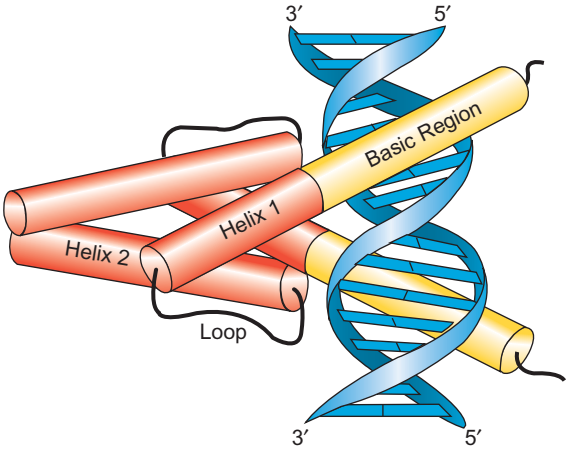


Figure 12.11 Crystal structure of the complex between the bHLH domain of MyoD and its DNA target. (a) Diagram with coiled ribbons representing α-helices. **(b)** Diagram with cylinders representing α-helices. (*Source:* Ma, P.C.M., M.A. Rould, H. Weintraub, and C.O. Palo, Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77 (6 May 1994) p. 453, f. 2a. Reprinted by permission of Elsevier Science.)

may require the extra interaction of the leucine zippers to ensure dimerization of the protein monomers.

SUMMARY The bZIP proteins dimerize through a leucine zipper, which puts the adjacent basic regions of each monomer in position to embrace the DNA target site like a pair of tongs. Similarly, the bHLH proteins dimerize through a helix-loop-helix motif, which allows the basic parts of each long helix to grasp the DNA target site, much as the bZIP proteins do. The bHLH and bHLH-ZIP domains bind to DNA in the same way, but the latter have extra dimerization potential due to their leucine zippers.

12.3 Independence of the Domains of Activators

We have now seen several examples of DNA-binding and transcription-activating domains in activators. These domains are separated physically on the proteins, they fold independently of each other to form distinct three-dimensional structures, and they operate independently of each other. Roger Brent and Mark Ptashne demonstrated this independence by creating a chimeric factor with the DNA-binding domain of one protein and the transcription-activating domain of the other. This hybrid protein functioned as an activator, with its specificity dictated by its DNA-binding domain.

Brent and Ptashne started with the genes for two proteins: GAL4 and LexA. We have already studied the DNA-binding and transcription-activating domains of GAL4; LexA is a prokaryotic repressor that binds to *lexA* operators and represses downstream genes in *E. coli* cells. It does not normally have a transcription-activating domain, because that is not its function. By cutting and recombining fragments of the two genes, Brent and Ptashne created a chimeric gene containing the coding regions for the transcription-activating domain of GAL4 and the DNA-binding domain of LexA. To assay the activity of the protein product of this gene, they introduced two plasmids into yeast cells. The first plasmid had the chimeric gene, which produced its hybrid product. The second contained a promoter responsive to GAL4 (either the *GAL1* or the *CYC1* promoter), linked to the *E. coli* β-galactosidase gene, which served as a reporter gene (Chapter 5). The more transcription from the GAL4-responsive promoter, the more β-galactosidase was produced. Therefore, by assaying for β-galactosidase, Brent and Ptashne could determine the transcription rate.

One more element was necessary to make this assay work: a binding site for the chimeric protein. The normal binding site for GAL4 is an upstream enhancer called UAS_G. However, this site would not be recognized by the chimeric protein, which has a LexA DNA-binding domain. To make the *GAL1* promoter responsive to activation, the investigators had to introduce a DNA target for the LexA DNA-binding domain. Therefore, they inserted a *lexA* operator in place of UAS_G. It is important to note that a *lexA* operator would not normally be found in a yeast cell; it was placed there just for the purpose of this experiment. Now the question is: Did the chimeric protein activate the *GAL1* gene?

The answer is yes, as Figure 12.12 demonstrates. The three test plasmids contained UAS_G, no target site, or the *lexA* operator. The activator was either LexA-GAL4, as we have discussed, or LexA (a negative control). With UAS_G present (Figure 12.12a), a great deal of β-galactosidase was made, regardless of which activator was present. This is because the yeast cells themselves make GAL4, which can activate via UAS_G. When no DNA target site was present (Figure 12.12b), no β-galactosidase could be made. Finally, when the *lexA* operator replaced UAS_G (Figure 12.12c), the LexA-GAL4 chimeric protein could activate β-galactosidase production over 500-fold. Thus, one can replace the

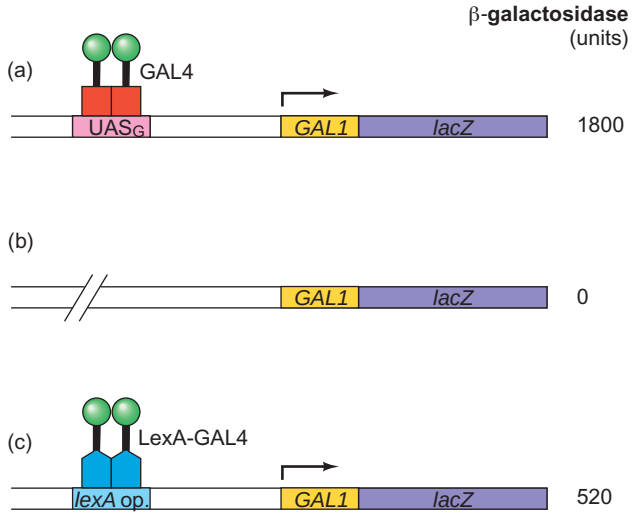


Figure 12.12 Activity of a chimeric transcription factor. Brent and Ptashne introduced two plasmids into yeast cells: (1) a plasmid encoding LexA-GAL4, a hybrid protein containing the transcription-activating domain of GAL4 (green) and the DNA-binding domain of LexA (blue); and (2) one of the test plasmid constructs shown in panels a–c. Each of the test plasmids had the *GAL1* promoter linked to a reporter gene (the *E. coli lacZ* gene). The chimeric protein LexA-GAL4 was used as the activator. The production of β-galactosidase (given at right) is a measure of promoter activity. **(a)** With a UAS_G element, transcription was very active and did not depend on the added transcription factor, because endogenous GAL4 could activate via UAS_G. **(b)** With no DNA target site, LexA-GAL4 could not activate, because it could not bind to the DNA near the *GAL1* promoter. **(c)** With the *lexA* operator, transcription was greatly stimulated by the LexA-GAL4 chimeric factor. The LexA DNA-binding domain could bind to the *lexA* operator, and the GAL4 transcription-activating domain could enhance transcription from the *GAL1* promoter.

DNA-binding domain of GAL4 with the DNA-binding domain of a completely unrelated protein, and produce a functional activator. This demonstrates that the transcription-activating and DNA-binding domains of GAL4 can operate quite independently.

SUMMARY The DNA-binding and transcription-activating domains of activator proteins are independent modules. We can make hybrid proteins with the DNA-binding domain of one protein and the transcription-activating domain of another, and show that the hybrid protein still functions as an activator.

12.4 Functions of Activators

In bacteria, the core RNA polymerase is incapable of initiating meaningful transcription, but the RNA polymerase holoenzyme can catalyze basal level transcription. Basal level transcription is frequently insufficient at weak promoters, so cells have activators to boost this basal transcription to higher levels by a process called **recruitment.** Recruitment leads to the tight binding of RNA polymerase holoenzyme to a promoter.

Eukaryotic activators also recruit RNA polymerase to promoters, but not as directly as prokaryotic activators. The eukaryotic activators stimulate binding of general transcription factors and RNA polymerase to a promoter. Figure 12.13 presents two hypotheses to explain this recruitment: (1) the general transcription factors cause a stepwise build-up of a preinitiation complex; or (2) the general transcription factors and other proteins are already bound to the polymerase in a complex called the **RNA polymerase II holoenzyme,** and the factors and polymerase are recruited together to the promoter. The truth may be a combination of the two hypotheses. In any event, it appears that direct contacts between general transcription factors and activators are necessary for recruitment. (However, as we will see later in this chapter, some activators require other proteins called coactivators to mediate the contact with the general transcription factors.) Which factors do the activators contact? The answer seems to be that many factors can be targets, but the one that was discovered first was TFIID.

Recruitment of TFIID

In 1990, Keith Stringer, James Ingles, and Jack Greenblatt performed a series of experiments to identify the factor that binds to the acidic transcription-activating domain of the herpesvirus transcription factor **VP16.** These workers expressed the VP16 transcription-activating domain as a fusion protein with the *Staphylococcus aureus* protein A, which binds tightly and specifically to immunoglobulin IgG.

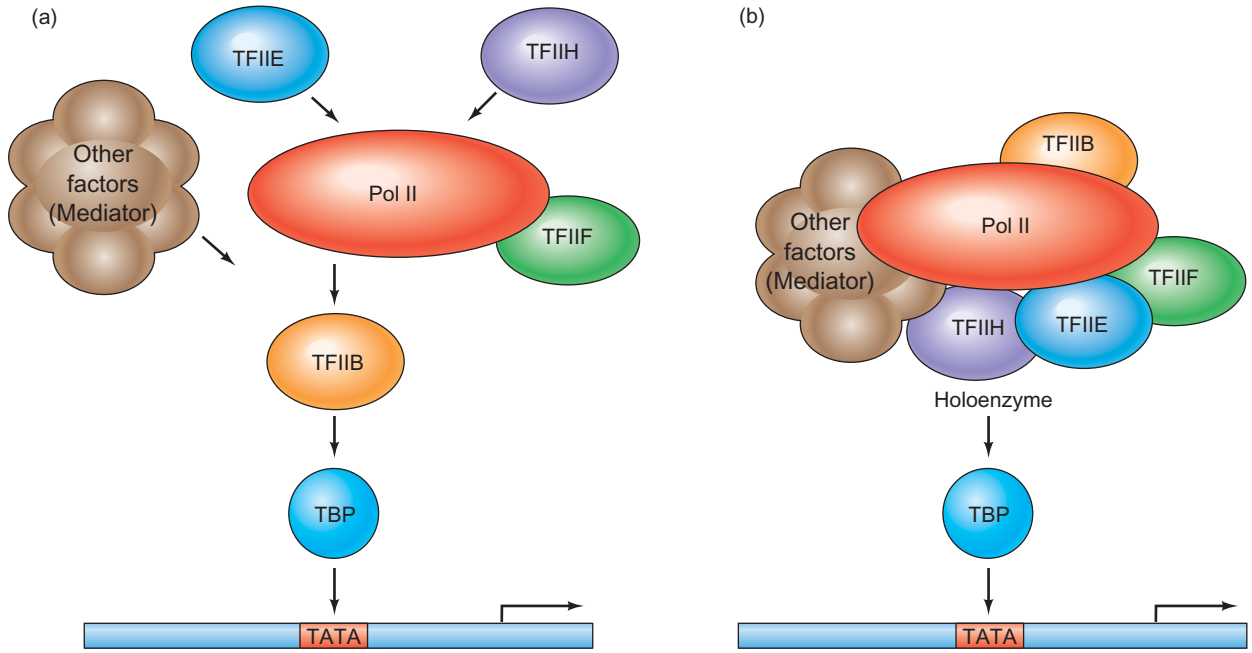


Figure 12.13 Two models for recruitment of yeast preinitiation complex components. (a) Traditional view of recruitment. This scheme calls for stepwise addition of components of the preinitiation complex, as occurs in vitro. **(b)** Recruitment of holoenzyme. Here, TBP binds first, then the holoenzyme binds to form the preinitiation complex. (*Source:* Adapted from Koleske, A.J. and R.A. Young, An RNA polymerase II holoenzyme responsive to activators. *Nature* 368:466, 1994.)

They immobilized the fusion protein (or protein A by itself) on an agarose IgG column and used these as affinity columns to "fish out" proteins that interact with the VP16-activating domain. To find out what proteins bind to the VP16-activating domain, they poured HeLa cell nuclear extracts through the columns containing either protein A by itself or the protein A/VP16-activating domain fusion protein. Then they used run-off transcription (Chapter 5) to assay various fractions for ability to transcribe the adenovirus major late locus accurately in vitro. They found that the flow-through from the protein A column still had abundant ability to support transcription, indicating no nonspecific binding of any essential factors to protein A. However, when they tested the flow-through from the protein A/VP16-activating domain column they found no transcription activity until they added back the proteins that bound to the column. Thus, some factor or factors essential for in vitro transcription bound to the VP16-activating domain.

Stringer and colleagues knew that TFIID was rate-limiting for transcription in their in vitro system, so they suspected that TFIID was the factor that bound to the affinity column. To find out, they depleted a nuclear extract of TFIID by heating it, then added back the material that bound to either the protein A column or the column containing the protein A/VP16-activating domain. Figure 12.14 shows that the material that bound to protein A by itself could not reconstitute the activity of a TFIID-depleted extract, but the material that bound to the protein A/VP16-activating domain could. This strongly suggested that TFIID binds to the VP16-activating domain.

To check this conclusion, Stringer and colleagues first showed that the material that bound to the VP16-activating domain column behaved just like TFIID on DEAE-cellulose ion-exchange chromatography. Then they assayed the material that bound to the VP16-activating domain column for the ability to substitute for TFIID in a template commitment experiment. In this experiment, they formed preinitiation complexes on one template, then added a second template to see whether it could also be transcribed. Under these experimental conditions, the commitment to transcribe the second template depended on TFIID. These workers found that the material that bound to the VP16-activating domain column could shift commitment to the second template, but the material that bound to the protein A column could not. These, and similar experiments performed with yeast nuclear extracts, provided convincing evidence that TFIID is the important target of the VP16 transcription-activating domain in this experimental system.

SUMMARY The acidic transcription-activating domain of the herpesvirus transcription factor VP16 binds to TFIID under affinity chromatography conditions.

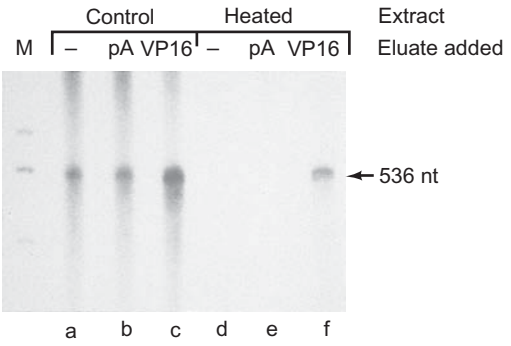


Figure 12.14 Evidence that an acidic activation domain binds TFIID. Stringer and colleagues fractionated a HeLa cell extract by affinity chromatography with a resin containing a fusion protein composed of protein A fused to the VP16-activating domain, or a resin containing just protein A. Then they eluted the proteins bound to each affinity column and tested them for ability to restore in vitro run-off transcription activity to an extract that had been heated to destroy TFIID specifically. Lanes a–c are controls in which the extract had not been heated. Because TFIID was still active, all lanes showed activity. Lanes d–f contained heated extract supplemented with: nothing (–), the eluate from the protein A column (pA), or the eluate from the column that contained the fusion protein composed of protein A and the transcription-activating domain of the VP16 protein (VP16). Only the eluate from the column containing the VP16 fusion protein could replace the missing TFIID and give an accurately initiated run-off transcript with the expected length (536 nt, denoted at right). Thus, TFIID must have bound to the VP16 transcription-activating domain in the affinity column. (*Source:* Stringer, K.F., C.J. Ingles, and J. Greenblatt, Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* 345 (1990) f. 2, p. 784. Copyright © Macmillan Magazines Ltd.)

Recruitment of the Holoenzyme

In Chapter 11 we learned that RNA polymerase II can be isolated from eukaryotic cells as a holoenzyme—a complex containing a subset of general transcription factors and other polypeptides. Much of our discussion so far has been based on the assumption that activators recruit general transcription factors one at a time to assemble the preinitiation complex. But it is also possible that activators recruit the holoenzyme as a unit, leaving only a few other proteins to be assembled at the promoter. In fact, there is good evidence that recruitment of the holoenzyme really does occur.

In 1994, Anthony Koleske and Richard Young isolated from yeast cells a holoenzyme that contained polymerase II, TFIIB, F, and H, and SRB2, 4, 5, and 6. They went on to demonstrate that this holoenzyme, when supplemented with TBP and TFIIE, could accurately transcribe a template bearing a *CYC1* promoter in vitro. Finally, they showed that the activator GAL4-VP16 could activate this transcription. Because the holoenzyme was provided intact, this last finding suggested that the activator recruited the intact holoenzyme to the promoter rather than building it up step by step on the promoter (recall Figure 12.13).

By 1998, investigators had purified holoenzymes from many different organisms, with varying protein compositions. Some contained most or all of the general transcription factors and many other proteins. Koleske and Young suggested the simplifying assumption that the yeast holoenzyme contains RNA polymerase II, a coactivator complex called Mediator, and all of the general transcription factors except TFIID and TFIIE. In principle, this holoenzyme could be recruited as a preformed unit, or piece by piece.

Evidence for Recruitment of the Holoenzyme as a Unit In 1995, Mark Ptashne and colleagues added another strong argument for the holoenzyme recruitment model. They reasoned as follows: If the holoenzyme is recuited as a unit, then interaction between *any* part of an activator (bound near a promoter) and *any* part of the holoenzyme should serve to recruit the holoenzyme to the promoter. This protein–protein interaction need not involve the normal transcription-activating domain of the activator, nor the activator's normal target on a general transcription factor. Instead, any contact between the activator and the holoenzyme should cause activation. On the other hand, if the preinitiation complex must be built up protein by protein, then an abnormal interaction between an activator and a seemingly unimportant member of the holoenzyme should not activate transcription.

Ptashne and colleagues took advantage of a chance observation to test these predictions. They had previously isolated a yeast mutant with a point mutation that changed a single amino acid in a holoenzyme protein (GAL11). They named this altered protein GAL11P (for potentiator) because it responded strongly to weak mutant versions of the activator GAL4. Using a combination of biochemical and genetic analysis, they found the source of the potentiation by GAL11P: The alteration in GAL11 caused this protein to bind to a region of the dimerization domain of GAL4, between amino acids 58 and 97. Because GAL11 (or GAL11P) is part of the holoenzyme, this novel association between GAL11P and GAL4 could recruit the holoenzyme to GAL4-responsive promoters, as illustrated in Figure 12.15. We call the association between GAL11P and GAL4 novel because the part of GAL11P involved is normally functionally inactive, and the part of GAL4 involved is in the dimerization domain, not the activation domain. It is highly unlikely that any association between these two protein regions occurs normally.

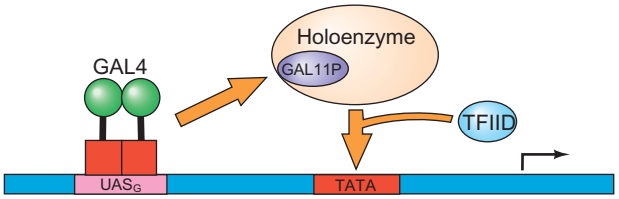


Figure 12.15 Model for recruitment of the GAL11P-containing holoenzyme by the dimerization domain of GAL4. The dimerization domain of GAL4 binds (orange arrow) to GAL11P (purple) in the holoenzyme. This causes the holoenzyme, along with TFIID, to bind to the promoter, activating the gene.

To test the hypothesis that the region of GAL4 between amino acids 58 and 97 is responsible for activation by GAL11P, Ptashne and colleagues performed the following experiment. Using gene-cloning techniques, they made a plasmid encoding a fusion protein containing the region between amino acids 58 and 97 of GAL4 and the LexA DNA-binding domain. They introduced this plasmid into yeast cells along with a plasmid encoding either GAL11 or GAL11P, and a plasmid bearing two binding sites for LexA upstream of a GAL1 promoter driving transcription of the *E. coli lacZ* reporter gene. Figure 12.16 summarizes this experiment and shows the results. The LexA-GAL4(58–97) protein is ineffective as an activator when wild-type GAL11 is in the holoenzyme (Figure 12.16a), but works well as an activator when GAL11P is in the holoenzyme (Figure 12.16b).

If activation is really due to interaction between LexA-GAL4(58–97) and GAL11P, we would predict that fusing the LexA DNA-binding domain to GAL11 would also cause activation, as illustrated in Figure 12.16c. In fact, this construct did cause activation, in accord with the hypothesis. Here, no novel interaction between LexA-GAL4 and GAL11P was required because LexA and GAL11 were already covalently joined.

The simplest explanation for these data is that activation, at least in this system, can operate by recruitment of the holoenzyme, rather than by recruitment of individual general transcription factors. It is possible, but not likely, that GAL11 is a special protein whose recruitment causes the stepwise assembly of a preinitiation complex. But it is much more likely that association between an activator and any component of the holoenzyme can recruit the holoenzyme and thereby cause activation. Ptashne and colleagues conceded that TFIID is an essential part of the preinitiation complex, but is apparently not part of the yeast holoenzyme. They proposed that TFIID might have bound to the promoter cooperatively with the holoenzyme in their experiments.

On the other hand, at least two lines of evidence suggest that the holoenzyme is not recruited as a whole. First, David Stillman and colleagues have performed kinetic studies of the binding of various factors to the *HO* promoter region in yeast. These studies showed that one part of the holoenzyme, Mediator, binds to the promoter earlier in G1 phase than does RNA polymerase II. Thus, the holoenzyme is certainly not binding as a complete unit, at least to this yeast promoter.

Second, Roger Kornberg and colleagues reasoned that, if the holenzyme binds as a unit to promoters, one should find all the components of the holoenzyme in roughly equal amounts in cells. They also knew that determining the concentrations of proteins in cells is tricky. One

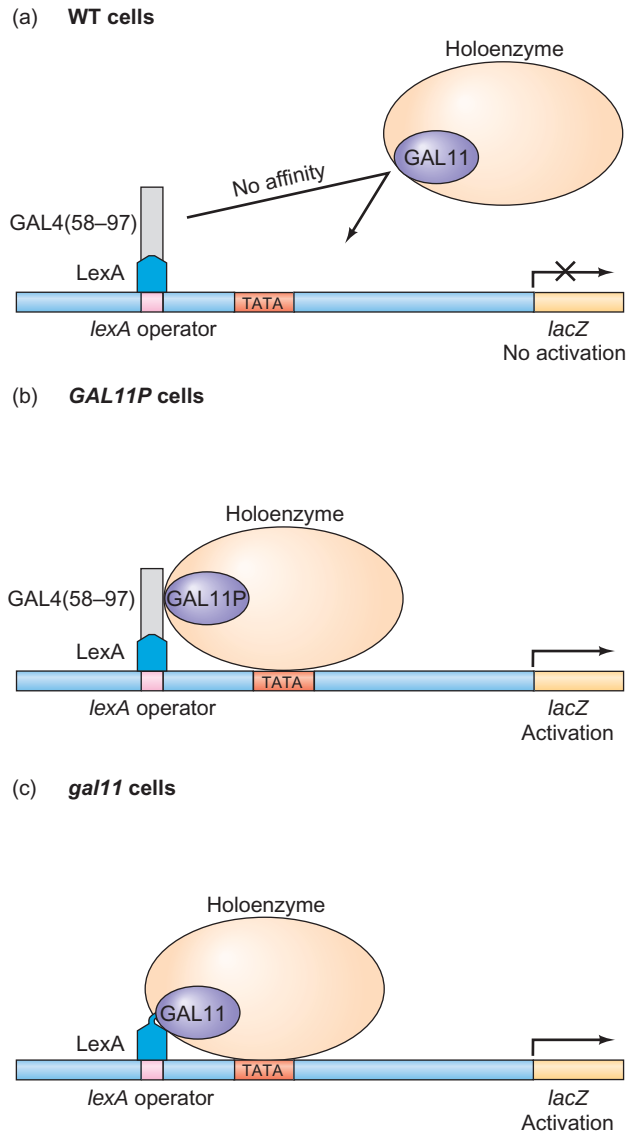


Figure 12.16 Activation by GAL11P and GAL11-LexA. Ptashne and colleagues transformed cells with a plasmid containing a *lexA* operator 50 bp upstream of a promoter driving transcription of a *lacZ* reporter gene, plus the following plasmids: **(a)** a plasmid encoding amino acids 58–97 of GAL4 coupled to the DNA-binding domain of LexA plus a plasmid encoding wild-type GAL11; **(b)** a plasmid encoding amino acids 58–97 of GAL4 coupled to the DNA-binding domain of LexA plus a plasmid encoding GAL11P; **(c)** a plasmid encoding GAL11 coupled to the DNA-binding domain of LexA. They assayed for production of the *lacZ* product, β-galactosidase. Results: **(a)** The GAL4(58–97) region did not interact with GAL11, so no activation occurred. **(b)** The GAL4(58–97) region bound to GAL11P, recruiting the holoenzyme to the promoter, so activation occurred. **(c)** The LexA-GAL11 fusion protein could bind to the *lexA* operator, recruiting the holoenzyme to the promoter, so activation occurred. (*Source:* Adapted from Barberis A., J. Pearlberg, N. Simkovich, S. Farrell, P. Resnagle, C. Bamdad, G. Sigal, and M. Ptashne, with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* 81:365, 1995.)

cannot do it by measuring mRNA levels because of wide variation in posttranscriptional events such as mRNA degradation and nuclear export. Indeed, concentrations of mRNAs and their respective protein products can deviate from expected values by up to 20- or 30-fold. One can separate proteins by two-dimensional gel electrophoresis and determine their concentrations by mass spectrometry (Chapter 24), but that method is not sensitive enough for proteins, such as transcription factors, found in very low concentrations in vivo.

So Kornberg and colleagues chose a method that combines high sensitivity and great accuracy. They began by using gene cloning techniques to attach "TAP" tags to the genes encoding seven different components of the polymerase II holoenzyme. These included RNA polymerase II, Mediator, and five general transcription factors. The TAP tag contains a region from *Staphylococcus* protein A (Chapter 4) that binds to antibodies of the IgG class. Thus, Kornberg and colleagues could dot-blot cell extracts from the yeast strains carrying genes for TAP-tagged proteins, then probe the blots with an antiperoxidase antibody. The TAP tag on a protein on the blot bound to the antibody, which in turn bound to peroxidase added later, which in turn converted a peroxidase substrate to a chemiluminescent product that could be detected photographically (Chapter 5).

The intensities of the bands on the film corresponded to the concentration of TAP-tagged proteins on the blots. With serial dilutions of each extract, these band intensities could be converted to concentrations of each protein per cell by comparing them with the results of a blot of known amounts of a standard, GST-TAP. Figure 12.17 shows sample results. It is clear from the wild-type lane with no TAP-tagged proteins that the background of this method is essentially zero, which is important for accuracy of quantification. It is also clear that there is considerably more RNA polymerase II than Med8, one of the subunits of Mediator. Quantification (Figure 12.17b) showed five to six times as much Rpb3 as any of the subunits of Mediator or of TFIIH. Table 12.1 presents a quantification of the amounts of TFIIF, TFIIE, TFIIB, and TFIID, in addition to the proteins considered in Figure 12.17. Again, RNA polymerase was more abundant than any of the other factors, but the four other general transcription factors were more abundant than either Mediator or TFIIH.

Because all of the components of the holoenzyme are *not* found in roughly equal amounts, it is unlikely that the holoenzyme binds to most promoters as a unit. It is still possible, though, that it is recruited to some promoters as a unit.

SUMMARY Activation, at least in certain promoters in yeast, appears to function by recruitment of the holoenzyme, rather than by recruitment of individual components of the holoenzyme one at a time. However, other evidence suggests that recruitment of the holoenzyme as a unit is not common.

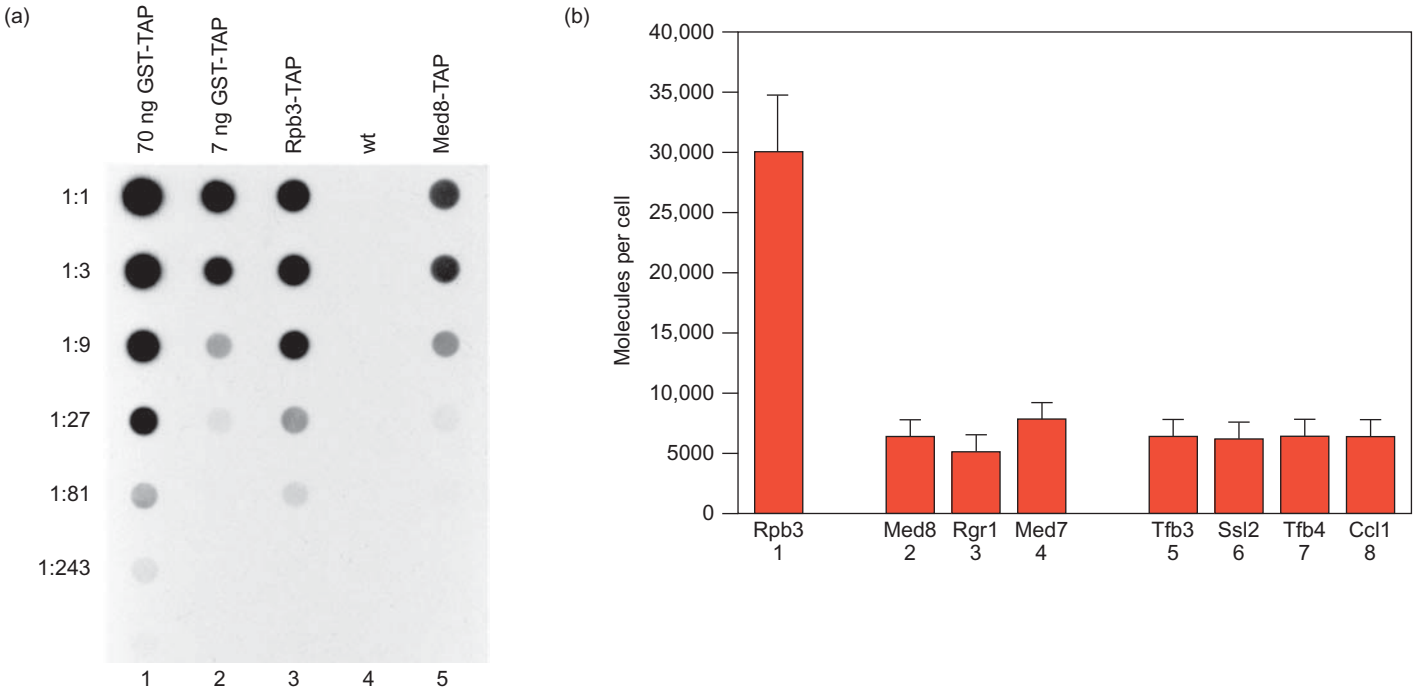


Figure 12.17 Determining the concentration of holoenzyme subunits by dot blotting. (a) Dot blot results. Kornberg and colleagues dot-blotted serial dilutions of extracts from cells bearing chimeric genes encoding holoenzyme subunits tagged with TAP sequences. The TAP sequences contained two *Staphylococcus* A protein sequences that bind to IgG immunoglobulins. The investigators reacted TAP sequences on the dot blot with an IgG immunoglobulin directed against peroxidase (rabbit antiperoxidase IgG). The IgG was in turn detected photographically with peroxidase and a substrate that becomes chemiluminescent on reaction with peroxidase. The dilutions are given at left. Columns 1 and 2 contained serial dilutions of two different amounts of GST-TAP, as given at top. Columns 3–5 contained serial dilutions of extracts from cells containing TAP-tagged Rpb3, wild-type cells with no TAP tags, and cells containing TAP-tagged Med8, respectively. **(b)** Cellular concentrations of Rpb3 (bar 1), three subunits of Mediator (bars 2–4), and four subunits of TFIIH (bars 5–8), determined by dot blotting. (*Source: Journal of Biological Chemistry* by Borggrefe et al. Copyright 2001 by Am. Soc. For Biochemistry & Molecular Biol. Reproduced with permission of Am. Soc. For Biochemistry & Molecular Biol. in the format Textbook via Copyright Clearance Center.)

Table 12.1 Number of Selected Protein Molecules per Yeast Cell

Protein	Copies per Cell
RNA polymerase II (Rpb3)	30,000
TFIIF (Tfg2)	24,000
TFIIE (Tfa2)	24,000
TFIIB (Sua7)	20,000
TFIID (TBP)	20,000
Mediator (Med8)	6000
TFIIH (Tfb3)	6000

Source: Borggrefe, T., R. Davis, A. Bareket-Samish, and R.D. Kornberg, Quantitation of the RNA polymerase II transcription machinery in yeast. *Journal of Biological Chemistry* 276 (2001): 47150–53, tII. Reprinted with permission.

12.5 Interaction Among Activators

We have seen several examples of crucial interactions among different types of transcription factors. Obviously, the general transcription factors must interact to form the preinitiation complex. But activators and general transcription factors also interact. For example, we have just learned that GAL4 and other activators interact with TFIID and other general transcription factor(s). In addition, activators usually interact with one another in activating a gene. This can occur in two ways: Individual factors can interact to form a protein dimer to facilitate binding to a single DNA target site. Alternatively, specific factors bound to different DNA target sites can collaborate in activating a gene.

Dimerization

We have already mentioned a number of different means of interaction between protein monomers in DNA-binding proteins. In Chapter 9 we discussed the helix-turn-helix proteins such as the λ repressor and observed that the interaction between the monomers of this protein place the recognition helices of the two monomers in just the right position to interact with two major grooves exactly one helical turn apart. The recognition helices are antiparallel to each other so they can recognize the two parts of a palindromic DNA target. Earlier in this chapter we discussed the coiled coil dimerization domains of the GAL4 protein and the similar leucine zippers of the bZIP proteins.

In Chapter 9 we discussed the advantage that a protein dimer has over a monomer in binding to DNA. This advantage can be summarized as follows: The affinity of binding between a protein and DNA varies with the square of the free energy of binding. Because the free energy depends on the number of protein–DNA contacts, doubling the contacts by using a protein dimer instead of a monomer quadruples the affinity between the protein and the DNA. This is significant because most activators have to operate at very low concentrations. The fact that the great majority of DNA-binding proteins are dimers is a testament to the advantage of this arrangement. We have seen that some activators, such as GAL4, form homodimers; others, such as the thyroid hormone receptor, form heterodimers.

SUMMARY Dimerization is a great advantage to an activator because it increases the affinity between the activator and its DNA target. Some activators form homodimers, but others function as heterodimers.

Action at a Distance

We have seen that both bacterial and eukaryotic enhancers can stimulate transcription, even though they are located some distance away from the promoters they control. How does this action at a distance occur? In Chapter 9 we learned that the evidence favors looping out of DNA in between the two remote sites to allow bacterial DNA-binding proteins to interact. We will see that this same scheme also seems to apply to eukaryotic enhancers.

Among the most reasonable hypotheses to explain the ability of enhancers to act at a distance are the following (Figure 12.18): (a) An activator binds to an enhancer and changes the topology, or shape, of the whole DNA duplex, perhaps by causing supercoiling. This in turn opens the promoter up to general transcription factors. (b) An activator binds to an enhancer and then slides along the DNA until it encounters the promoter, where it can activate transcription by virtue of its direct contact with the promoter DNA. (c) An activator binds to an enhancer and, by looping out DNA in between, interacts with proteins at the promoter, stimulating transcription. (d) An activator binds to an enhancer and a downstream segment of DNA to form a DNA loop. By enlarging this loop, the protein tracks toward the promoter. When it reaches the promoter, it interacts with proteins there to stimulate transcription.

Notice that the first two of these models demand that the two elements, enhancer and promoter, be on the same DNA molecule. A change in topology of one DNA molecule cannot influence transcription on a second, and an activator cannot bind to an enhancer on one DNA and slide onto a second molecule that contains the promoter. On the other hand, the third model simply requires that the enhancer and promoter be relatively near each other, not necessarily on the same molecule. This is because the essence of the looping model is not the looping itself, but the interaction between the proteins bound to remote sites. In principle, this would work just as well if the proteins were bound to two sites on different DNA molecules, as long as the molecules were tethered together somehow so they would not float apart and prevent interactions between the bound proteins. Figure 12.19 shows how this might happen.

Thus, if we could arrange to put an enhancer on one DNA molecule and a promoter on another, and get the two molecules to link together in a **catenane,** (circles linked as in a chain) we could test the hypotheses. If the enhancer still functioned, we could eliminate the first two. Marietta Dunaway and Peter Dröge did just that. They constructed a plasmid with the *Xenopus laevis* rRNA promoter plus an rRNA minigene on one side and the rRNA enhancer on the other, with the λ phage integration sites, *attP* and *attB,* in between. These are targets of site-specific recombination, so placing them on the same molecule and allowing recombination produces a catenane, as illustrated in Figure 12.19.

Finally, these workers injected combinations of plasmids into *Xenopus* oocytes and measured their transcription by quantitative S1 mapping. The injected plasmids were the catenane, the unrecombined plasmid containing both enhancer and promoter, or two separate plasmids, each containing either the enhancer or promoter. In quantitative S1 mapping, a reference plasmid is needed to correct for the variations among oocytes. In this case, the reference plasmid contained an rRNA minigene (called ψ52) with a 52-bp insert, whereas the rRNA minigenes of the test plasmids (called ψ40) all contained a 40-bp insert. Dunaway and Dröge included probes for both these minigenes in their assay, so we expect to see two signals, 12 nt apart, if both genes are transcribed. We are most interested in the *ratio* of these two signals, which tells us how well each test plasmid is transcribed relative to the reference plasmid, which should behave the same in each case.

Figure 12.20a shows the test plasmid results in the lanes marked "a" and the reference plasmid results in the lanes marked "b." The plasmids used to produce the transcripts in each lane are pictured in panel (b). Note that the same plasmids were used in both lane a and lane b of each set in panel (a). Only the probes were different. These were the results: Lanes 1 show that when the plasmid contained the promoter alone, the test plasmid signal was weaker than the reference plasmid signal. That is because the test probe was less radioactive than the reference probe. Lanes 2 demonstrate that the enhancer adjacent to the promoter (its normal position) greatly enhanced transcription in the test plasmid—its signal was much stronger than the reference plasmid signal. Lanes 3 show that the enhancer still worked, though not quite as well, when placed opposite the promoter on the plasmid. Lanes 4 are the most important. They show that the enhancer still worked when it was on a separate plasmid that formed a catenane with the

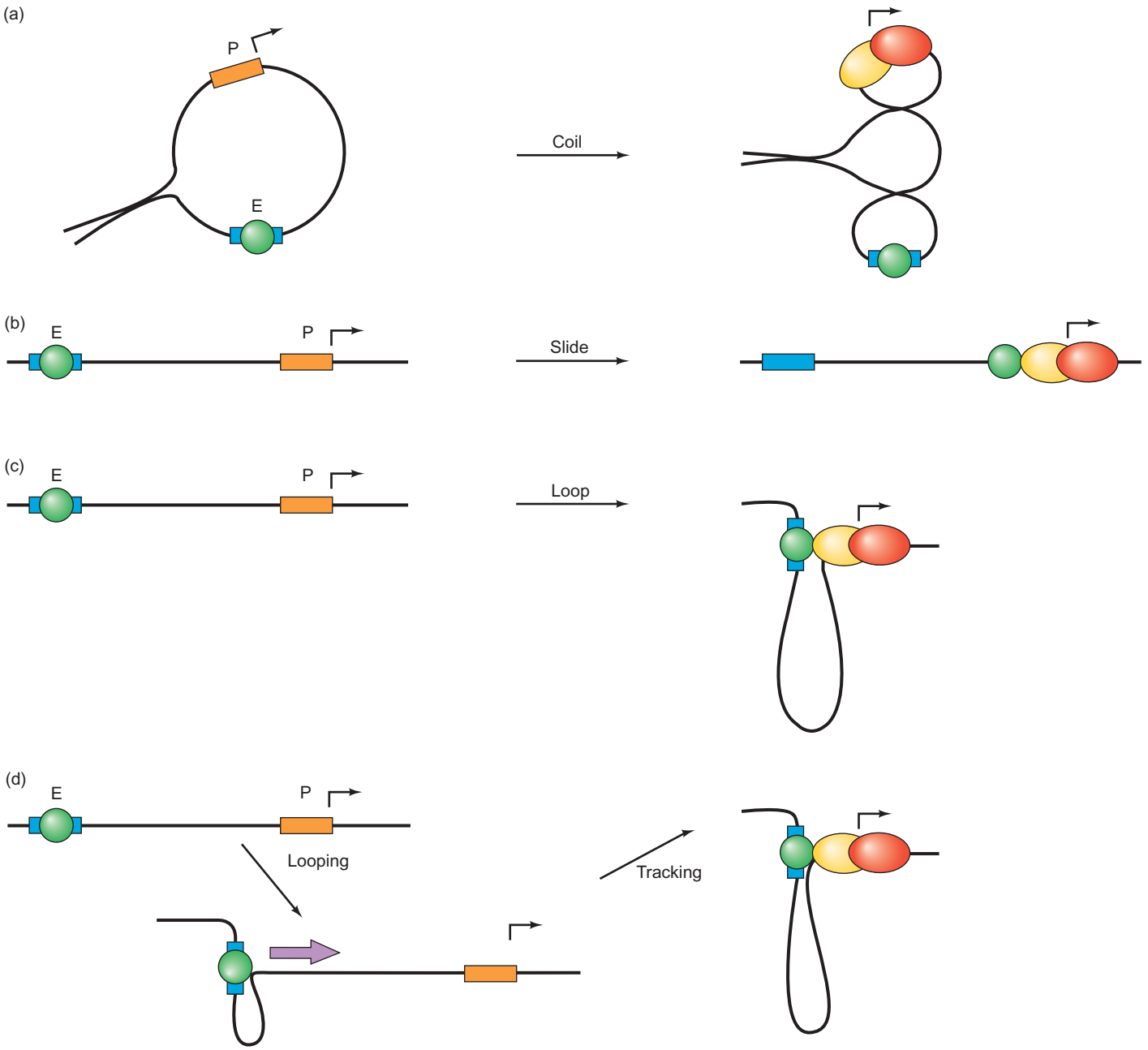


Figure 12.18 Four hypotheses of enhancer action. (a) Change in topology. The enhancer (E, blue) and promoter (P, orange) are both located on a loop of DNA. Binding of a gene-specific transcription factor (green) to the enhancer causes supercoiling that facilitates binding of general transcription factors (yellow) and polymerase (red) to the promoter. **(b)** Sliding. A transcription factor binds to the enhancer and slides down the DNA to the promoter, where it facilitates binding of general transcription factors and polymerase. **(c)** Looping. A transcription factor binds to the enhancer and, by looping out the DNA in between, binds to and facilitates the binding of general transcription factors and polymerase to the promoter. **(d)** Facilitated tracking. A transcription factor binds to the enhancer and causes a short DNA segment to loop out downstream. Increasing the size of this loop allows the factor to track along the DNA until it reaches the promoter, where it can facilitate the binding of general transcription factors and RNA polymerase.

plasmid containing the promoter. Lanes 5 verify that the enhancer did not work if it was on a separate plasmid *not* linked in a catenane with the promoter plasmid. Finally, lanes 6 show that the enhancement observed in lanes 4 was not due to a small amount of contamination by unrecombined plasmid. In lanes 6, the investigators added 5% of such a plasmid and observed no significant increase in the test plasmid signal.

These results lead to the following conclusion about enhancer function: The enhancer does not need to be on the same DNA with the promoter, but it does need to be able to approach the promoter, so the proteins bound to

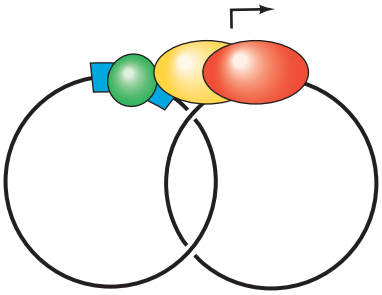


Figure 12.19 Interaction between enhancer and promoter on two plasmids linked in a catenane. Hypothetical interaction between an activator (green) bound to an enhancer (blue) on one plasmid, and general transcription factors (yellow) and RNA polymerase (red) bound to the promoter (not visible beneath the bent arrow) in the other plasmid of the catenane.

enhancer and promoter can interact. This is difficult to reconcile with models involving supercoiling or sliding (Figure 12.18a and b), but is consistent with the DNA looping and facilitated tracking models (Figure 12.18c and d). In the catenane, no looping or tracking is required because the enhancer and promoter are on different DNA molecules; instead, protein–protein interactions can occur without looping, as illustrated in Figure 12.19a.

If enhancer action requires DNA looping, then we should be able to observe it directly, using appropriate tools. A technique called **chromosome conformation capture** (3C) provides just such a tool. This method, illustrated in Figure 12.21, is designed to test whether two remote DNA regions, such as an enhancer and a promoter, are brought together—by interactions between DNA-binding

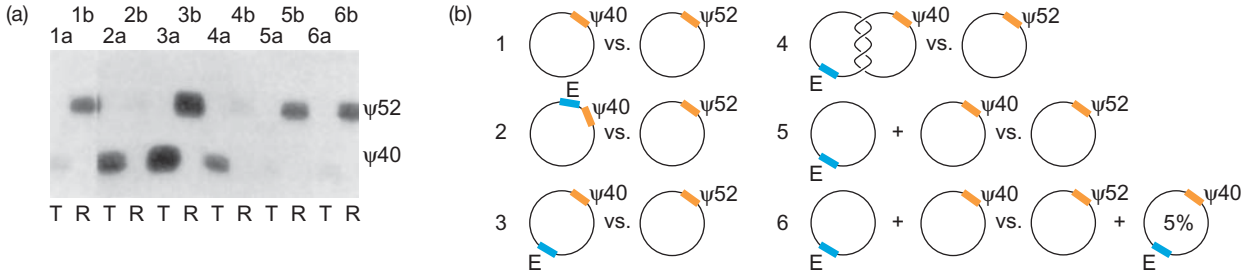


Figure 12.20 Results of the catenane experiment. Dunaway and Dröge injected mixtures of plasmids into *Xenopus* oocytes and measured transcription rates by quantitative S1 mapping. They injected a test plasmid and a reference plasmid in each experiment and assayed for transcription of each with separate probes. **(a)** Experimental results. The results of the test (T) and reference (R) assays are given in lanes a and b, respectively, of each experiment. The plasmids injected in each experiment are given in panel **(b).** For example, the plasmids used in the experiments in lanes 1a and 1b are labeled 1. The plasmids on the left, labeled Ψ40 (or Ψ40 plus another plasmid), are the test plasmids. The ones on the right, labeled Ψ52, are the reference plasmids. The 40 and 52 in these names denote the size inserts each has to distinguish it from the other. Both plasmids were injected and then assayed with the test probe (lane 1a) or the reference probe (lane 1b). Lanes 4a and 4b demonstrate that transcription of the catenane with the enhancer on one plasmid and the promoter on the other is enhanced relative to transcription of the plasmid containing just the promoter (lanes 1a and 1b). This is evident in the much higher ratio of the signals in lanes 4a and 4b relative to the ratio of the signals in lanes 1a and 1b. (*Source:* Adapted from Dunaway M. and P. Dröge, Transactivation of the *Xenopus* rRNA gene promoter by its enhancer. *Nature* 341 (19 Oct 1989) p. 658, f. 2a. Copyright © Macmillan Magazines Ltd.)

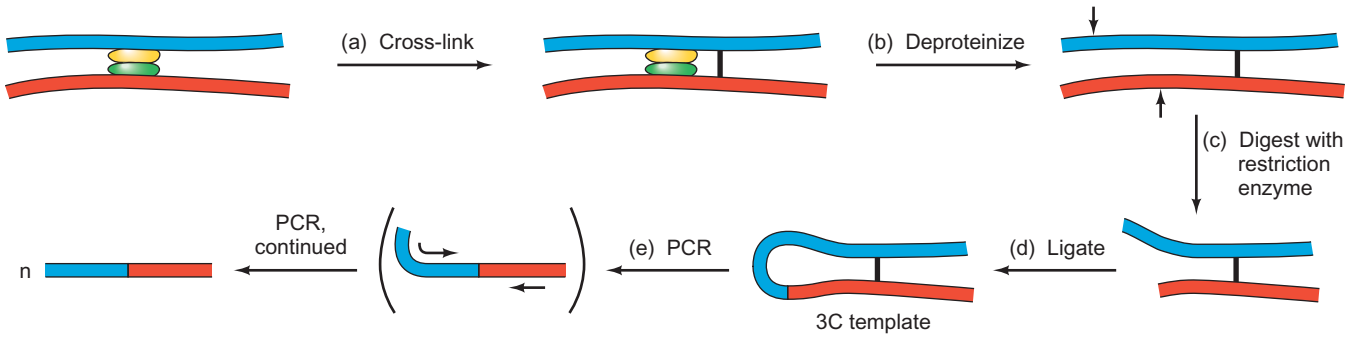


Figure 12.21 Chromatin conformation capture (3C). (a) Begin with chromatin in which you believe two sites are brought together by interaction between two DNA-binding proteins (green and yellow). The two segments of chromosome (red and blue) can be on separate chromosomes, or the same chromosome. Cross-link the two separate chromosome segments with formaldehyde. **(b)** Deproteinize the chromatin. **(c)** Digest the DNA with a restriction enzyme. Arrows show two restriction sites. **(d)** Ligate the nearby DNA ends under conditions (low DNA concentration) in which intramolecular ligation is favored. This yields the 3C template. **(e)** PCR on the 3C template with primers indicated by the short arrows yields a significant amount of PCR product, showing that the two chromosome segments represented by the primers are probably close together in this chromatin.

BOX 12.1

Genomic Imprinting

Because most eukaryotes are diploid organisms, you would probably predict that it doesn't matter which allele of any gene pair came from the mother and which came from the father. In most cases, you would be right, but there are important exceptions. The first evidence for one very important class of exceptions came from studies with mouse eggs just after fertilization, in which the maternal and paternal nuclei had not yet fused. At this stage, the maternal nucleus can be removed and replaced with a second paternal nucleus. Similarly, the paternal nucleus can be removed and replaced with a second maternal nucleus. In either case, the embryo will have chromosomes contributed by only one parent. In principle, that should not have made a big difference, because the parental mice were from an inbred strain in which all the individuals are genetically identical (except, of course, for the XY versus XX difference between males and females).

In fact, however, it made a tremendous difference. All of these embryos died during development, most at a very early stage. Those that made it the longest before dying showed an interesting difference, depending on whether their genes came from the mother or the father. Those with genes derived only from the mother had few abnormalities in the embryo itself, but had abnormal and stunted placentas and yolk sacs. Embryos with genes derived only from the father were small and poorly formed, but had relatively normal placentas and yolk sacs. How can we account for this difference if the genes contributed by the mother and father are identical? One explanation for this phenomenon is that the genes—that is, the base sequences of the genes—are identical, but they are somehow modified, or imprinted, differently in males and females.

Bruce Cattanach provided more evidence for imprinting with his studies on mice with fused chromosomes. For example, in some mice, chromosome 11 is fused, so it cannot separate during mitosis or meiosis. This means that some gametes produced by such a mouse will have two copies of chromosome 11, while some will have none. These mice made it possible for Cattanach to produce offspring with both chromosomes 11 from the father (using sperm with a double dose of chromosome 11 and eggs with no chromosome 11, or both from the mother (by reversing the procedure). Again, if the parental source of the chromosome did not matter, these offspring should have been normal. But they were not. In cases where both chromosomes came from the mother, the pups were abnormally small; if both chromosomes came from the father, the pups were giants.

Furthermore, these experiments demonstrated that the imprint is erased at each generation. That is, a runty male mouse whose chromosomes 11 came from his mother generally would produce normal-size offspring himself. The production of male gametes somehow erased the maternal imprint.

Genomic imprinting also occurs in humans, occasionally with tragic results. Inheritance of a deleted chromosome 15 from the father is associated with Prader-Willi syndrome, in which the patient is typically mentally impaired, short, and obese, because of an uncontrollable appetite. The lack of a particular part of the paternal copy of chromosome 15 is important because the gene associated with Prader-Willi syndrome is imprinted, and therefore inactivated, on the maternal chromosome 15. Thus, deletion of the paternal allele, and imprinting of the maternal allele, leaves no functioning copy of the gene. By contrast, inheritance of a deleted chromosome 15 from the mother is connected with Angelman syndrome, characterized by a large mouth and abnormally red cheeks, as well as by severe mental impairment, with inappropriate laughter and jerky movements. The lack of a particular part of the maternal

proteins, for example. First, chromatin with suspected DNA looping is fixed with formaldehyde to form covalent bonds between chromatin regions that are in close contact. (Chromatin is the natural state of DNA within a eukaryotic cell. It consists of DNA bound to an approximately equal mass of protein (Chapter 13). Next, the chromatin is deproteinized and digested with a restriction enzyme (Chapter 4). Next, the free DNA ends are ligated together to form a so-called 3C template. If two formerly remote regions of chromatin are in contact with each other, they will be ligated together in the 3C template, and PCR primers specific for these two regions will produce a relatively short PCR product. The more prevalent this product, the more often the two chromatin regions are in contact. This method can be used to detect either intra- or interchromosomal interactions.

Karl Pfeifer and colleagues exploited the 3C method to demonstrate interaction between an enhancer and a promoter. They focused on the mouse ***Igf2/H19* locus** (Figure 12.22a). The *Igf2* gene, driven by three promoters, spaced 2 kb apart, encodes IGF2 (**interferon-like growth factor 2**), and *H19* encodes a noncoding RNA. Interestingly, the *Igf2* gene on the male chromosome is turned on, but the homologous gene on the female chromosome is silenced. Conversely, the *H19* gene on the female chromosome is on, but the homologous gene on the male chromosome is off. This chromosome-specific behavior is explained by **imprinting,** which is established during gametogenesis by methylation of the **imprinting control region** (***ICR***). Box 12.1 gives further insight into the biology of imprinting, and this locus in particular. Later in this chapter, we will learn more about the mechanism of imprinting.

copy of chromosome 15 is important because the gene, or genes, associated with Angelman syndrome are imprinted, and therefore inactivated, on the paternal chromosome. Thus, deletion of the maternal copies, and imprinting of the paternal copies, leaves no functioning copies of these genes.

How can the DNA be modified in a reversible way so the imprint can be erased? The evidence points to DNA methylation. First, experiments show that genes derived from males and females are methylated differently, and this methylation correlates with gene activity. In general, methylated genes are found in females, and the methylated genes are inactivated. (However, note that in the *Igf2* example in the main text, it is an insulator that gets methylated in male mice, and this allows *Igf2* expression, whereas the unmethylated insulator in females blocks *Igf2* expression.)

Furthermore, methylation can be reversed. Philip Leder and colleagues used transgenic mice (Chapter 5) to follow the methylated state of a transgene as it moves through gametogenesis (the production of sperm or eggs) and into the developing embryo. These experiments revealed that the methyl groups on the transgene are removed in the early stages of gametogenesis in both males and females. The developing egg then establishes the maternal methylation pattern before the oocyte is completely mature. In the male, some methylation occurs during sperm development, but this methylation pattern is further modified in the developing embryo. Thus, methylation has all the characteristics we expect in an imprinting mechanism: It occurs differently in male and female gametes; it is correlated with gene activity; and it is erased after each generation.

Do any benefits derive from genomic imprinting, or is it just another cause of genetic disorders? David Haig has cited an imprinting example that he believes has evolved in response to environmental demands: The insulin-like growth factor (IGF-2), and its receptor in the mouse. The growth factor tends to make baby mice bigger, but it must interact with its receptor (the type-1 IGF receptor) in order to do so. To complicate the problem, mice have an alternate receptor (a type-2 receptor) that binds IGF-2 but does not pass the growth-promoting signal along. Thus, expression of the *Igf2* gene in developing mice will produce bigger offspring, but expression of the type-2 receptor will sop up the IGF-2 and keep it away from the type-1 receptor, and therefore produce smaller offspring.

Haig points to an inherent biological conflict between the interests of the mother and those of the father of a baby mammal. If the benefits to the mother and father are viewed simply in terms of getting their own genes passed on to their offspring, then the father should favor large offspring, and the mother should favor small ones. The reason is that a large baby is more likely to survive and therefore perpetuate the father's genes. On the other hand, a large baby saps the mother's strength and leaves her fewer resources to provide to other offspring, which could be sired by a different father, but still would perpetuate her genes. This is a coldhearted way of looking at parenthood, but it is the sort of thing that can influence evolution.

Viewed in this context, it is very interesting that imprinting of male and female gametes in the mouse dictate that the *Igf2* gene provided by a mother mouse is repressed, while that provided by the father is active. On the other hand, the type-2 IGF receptor gene from the father is turned off, whereas that from the mother is active. Both of these phenomena fit with the premise that a male should favor large offspring and a female should favor small ones. We seem to have a battle of the sexes going on at the molecular level, but neither side is winning, because the strategies of each side are canceled by those of the other!

The *Igf2/H19* locus also contains two enhancers, one of which is active in endodermal cells, and the other in mesodermal cells. These enhancers can stimulate transcription of both the *Igf2* and *H19* genes. Notice that the ICR lies between the enhancers and the *Igf2* promoters, but not between the enhancers and the *H19* promoter. This location enables the ICR to function as an insulator to shield the *Igf2* promoters from the stimulatory effect of the enhancers, but only on the maternal chromosome. We will learn about insulator activity later in this chapter; for now, it is sufficient to know that the *Igf2* gene is active only on the paternal chromosome.

The imprinted nature of the *Igf2* locus allowed Pfeifer and colleagues to look at DNA looping between enhancers and promoters on active (paternal) and inactive (maternal) chromosomes in the same cells. If the looping model of enhancer action is correct, such looping would be observed only on the paternal chromosomes—and that is what happened.

To distinguish between maternal and paternal chromosomes in the 3C experiments, Pfeifer and colleagues bred mice that had *Igf2* loci from two different mouse species, as follows: They intercrossed FVB mice (*Mus domesticus*) with Cast7 mice, which are just like FVB mice, but have the distal part of chromosome 7, including the *Igf2* locus, derived from another mouse species (*Mus castaneus*). The *Igf2* loci of the two mouse species differ in several restriction sites, so cleavage with certain restriction enzymes yields different-size restriction fragments from DNAs of the two species. These variations are called **restriction fragment length polymorphisms (RFLPs,** Chapter 24), and can be used to determine whether a PCR product in a 3C experiment comes from the maternal or paternal chromosome.

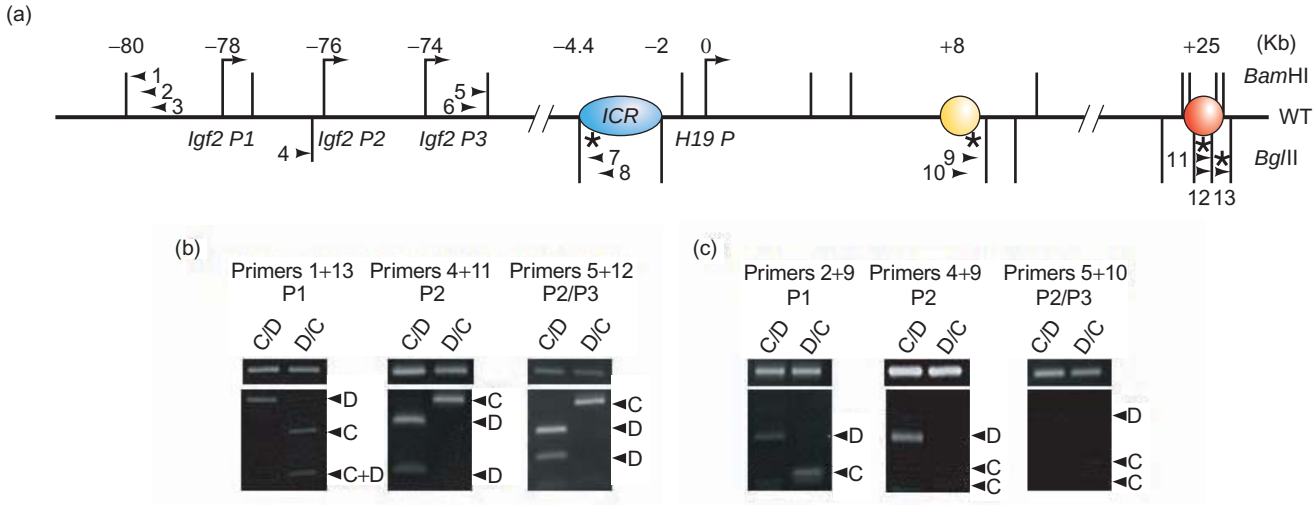


Figure 12.22 Association of chromatin elements in the mouse *Ifg2* locus. (a) Map of the wild-type locus. The whole locus is just over 100 kb long, as indicated at top. The three *Igf2* promoters are indicated near positions −78, −76, and −74, and the *H19* promoter is indicated at position 0. The ICR is in blue and the endodermal and mesodermal enhancers are in yellow and red, respectively. The vertical bars above and below the DNA represent *Bam*HI and *Bgl*II sites, respectively. Asterisks indicate *Bgl*II RFLPs that distinguish between *M. domesticus* and *M. castaneus* DNAs. Short arrows represent PCR primers used in the 3C analysis. Note that these primers always point toward the nearby restriction site. Thus, they are in position to create a short PCR product whenever two remote sections of DNA are cut with the corresponding restriction enzyme and then ligated together. (**b-c**) 3C analysis of long-range interactions in (**a**) mouse fetal muscle (mesodermal) cells and (**b**) fetal liver (endodermal) cells, respectively, using the indicated primers. The source of the embryo chromosomes (*M. domesticus* [DI or *M. castaneus* [C]) is shown at top of each panel, with the maternal chromosome first. The upper panels in each case show the PCR product of the 3C analysis. The lower panels show the RFLP analysis on the PCR products. Arrowheads labeled C or D point to RFLP bands that are characteristic of *M. castaneus* or *M. domesticus,* respectively. C+D denotes an RFLP band resulting from comigration of bands from both mouse species. (*Source:* Yoon et al, Analysis of the H19/*ICR*. *Molecular and Cellular Biology,* May 2007, pp. 3499–3510, Vol. 27, No. 9. Copyright © 2007 American Society for Microbiology.)

Figure 12.22b and c show the 3C results in fetal muscle (mesodermal) cells and fetal liver (endodermal) cells, respectively. The top part of each panel contains the 3C PCR product, and the bottom part contains the results of RFLP analysis to identify the maternal or paternal origin of each PCR product. The C/D and D/C designations at the top refer to the *M. castaneus* or *M. domesticus Igf2* locus, with the maternal allele always presented first. Thus, C/D mice had the *M. cataneus Igf2* locus on the maternal chromosome and the *M. domesticus Igf2* locus on the paternal chromosome. The C and D designations beside the gels show RFLP bands corresponding to *M. castaneus* and *M. domesticus,* respectively. Note that the 3C PCR products always derived from the paternal chromosome. For example, in the first lane in the first gel in Figure 12.22b, the paternal chromosome was from *M. domesticus,* and the RFLP analysis identified the PCR product as coming from *M. domesticus* (D). On the other hand, in the second lane in the first gel, the paternal chromosome was from *M. castaneus,* and the RFLP analysis showed that the PCR product came from *M. castaneus* (C). This demonstrated that the enhancer and promoters are brought together by DNA looping only on the paternal chromosome, where the *Igf2* gene is active.

Pfeifer and colleagues chose the primers to show linkages between each of the three *Igf2* promoters and the appropriate enhancer. Thus, in muscle cells, DNA looping brought each of the promoters (defined by primers 1, 4, and 5, respectively), close to the mesodermal enhancer (the one on the far right in Figure 12.22a, and defined by primers 11, 12, and 13). On the other hand, in liver cells, DNA looping brought the promoters and the endodermal enhancer (defined by primers 9 and 10) together. Thus, the 3C technique demonstrates that tissue-appropriate enhancers and promoters are brought together, presumably by DNA looping.

SUMMARY The essence of enhancer function—protein–protein interaction between activators bound to the enhancers, and general transcription factors and RNA polymerase bound to the promoter—seems in many cases to be mediated by looping out the DNA in between. This can also account for the effects of multiple enhancers on gene transcription, at least in theory. DNA looping could bring the activators bound to each enhancer close to the promoter where they could stimulate transcription, perhaps in a cooperative way.

Transcription Factories

The notion of DNA loops discussed in the previous section is consistent with the concept of **transcription factories**—discrete nuclear sites where transcription of multiple genes

occurs: If two or more active genes on the same chromosome are clustered in the same transcription factory, this would naturally form DNA loops between them. Thus, the existence of transcription factories implies the existence of DNA loops in eukaryotic nuclei. During the 1990s, several research groups provided evidence for the existence of these transcription factories. This concept raises at least two interesting questions: (1) How many transcription factories exist in a nucleus? (2) How many polymerases are active in a transcription factory?

To count the number of transcription factories, Peter Cook and colleagues performed the following experiment in 1998. They labeled growing RNA chains in HeLa cells with bromouridine (BrU). They followed this BrU labeling in vivo by permeabilizing the cells and further labeling growing RNA chains in vitro with biotin-CTP. The labeled RNA could then be detected with primary antibodies against either BrU or biotin, and secondary antibodies or protein A labeled with gold particles. BrU labeling was detected with 9-nm gold particles, and biotin labeling was detected with 5-nm particles. Figure 12.23a shows the results of labeling with BrU at low magnification, and Figure 12.23b shows the results of labeling with both BrU and biotin at higher power. Note that transcription does not occur uniformly across the nucleus, but is concentrated into patches, most of which contain more than one growing RNA chain.

The purpose of the in vitro labeling with biotin is to control for migration of finished RNAs away from their site of synthesis. If RNAs do this in groups, these would appear just like transcription factories and the number of apparent factories would therefore be inflated. But labeling in vitro does not allow for RNA chains to be finished and leave their sites of synthesis, so in vitro-labeled RNAs (small gold particles) should represent real transcription factories. Cook and colleagues found a high level of correspondence between in vivo- and in vitro–labeled clusters, as long as the in vivo labeling times were kept short (2.5 min). That is, large gold particles were found in the same clusters with small gold particles about 85% of the time. With longer in vivo labeling times (10 min or more), many BrU-labeled clusters were not associated with biotin-labeled clusters, and were therefore probably not transcription factories.

Do the clusters really represent sites of transcription? If so, we would expect the number of particles to increase with time, as more polymerases initiate RNA chains. Figure 12.23c shows that the number of particles in clusters does indeed increase with time, while the number of single particles does not. Thus, transcription is associated with the clusters, not the single particles.

On average, Cook and colleagues found one cluster per μm^2: in their nuclear sections. Knowing the total nucleoplasmic volume, this allowed them to calculate that there are about 5500 nucleoplasmic transcription factories with active polymerases II and III per cell. Extending preinitiated RNA chains in vitro with labeled UTP in the presence and

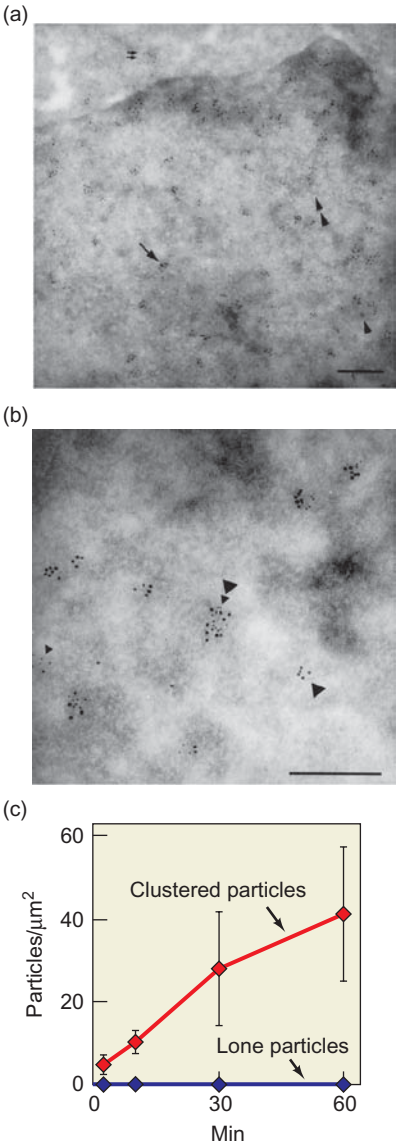


Figure 12.23 Detecting transcription factories. (a) Low-magnification view. Cook and colleagues labeled growing RNA chains in HeLa cells with BrU and detected the label by indirect immunostaining with 9-nm gold particles. They found most of the labeled RNA in clusters (arrow). Most of these clusters represent transcription factories, but some represent sites of RNA processing, or even mature RNAs in the cytoplasm (two small arrows). Weak label was found in interchromatin clusters (double arrowhead). No label was found in perichromatin clusters (single arrowhead). **(b)** High-magnification view. Cook and colleagues labeled nascent RNA with BrU in vivo and then extended these growing RNAs in vitro and labeled them with biotin-CTP. They detected BrU- and biotin-labeled RNAs by indirect immunostaining with 9-nm and 5-nm gold particles, respectively. They found most gold particles in clusters. Large and small arrowheads point to clusters with large and small gold particles, respectively. Most clusters contained both sizes of particles. **(c)** Clustered particles correspond to transcription sites. Cook and colleagues grew cells for various times in medium containing BrU, then detected BrU-RNA by immunostaining with 9-nm gold particles. (*Source:* Jackson et al, Numbers and Organization of RNA Polymerases, Nascent Transcripts, and Transcription Units in HeLa Nuclei. *Molecular Biology of the Cell* Vol. 9, 1523–1536, June 1998. Copyright © 1998 by The American Society for Cell Biology.)

absence of α-amanitin gave Cook and colleagues an estimate of the total amount of RNA synthesized during the in vitro labeling period. Knowing the approximate length each RNA chain would grow during the labeling period, these workers could estimate the number of growing RNA chains, and therefore the number of active polymerases. They calculated that each cell contained about 75,000 active RNA polymerases II and III. Thus, given that there are about 5500 transcription factories per cell, there are about 75,000/5500, or about 14 active polymerases II and III per transcription factory.

SUMMARY Transcription appears to be concentrated in transcription factories within the nucleus, where an average of about 14 polymerases II and III are active. The existence of transcription factories implies the existence of DNA loops between genes being transcribed in the same factory.

Complex Enhancers

Many genes have more than one activator-binding site, so they can respond to multiple stimuli. For example, the metallothionine gene, which codes for a protein that apparently helps eukaryotes cope with poisoning by heavy metals, can be turned on by several different agents, as illustrated in Figure 12.24. Thus, each of the activators that bind at these sites must be able to interact with the preinitiation complex assembling at the promoter, presumably by looping out any intervening DNA.

The finding that multiple activator-binding sites can control a given gene is changing our definition of the word "enhancer." It was originally defined as a nonpromoter DNA element that, together with at least one enhancer-binding protein, could stimulate transcription of a nearby gene. Thus, the control region of the metallothionine gene upstream of the TATA box in Figure 12.24 was considered to contain many enhancers. But the definition has evolved toward a concept that embraces an entire contiguous control region outside the promoter itself. Thus, the entire control region of the metallothionine gene can be considered an enhancer, and the BLE, for example, is only one element of the whole enhancer. Even using the newer definition, we can still say that some genes are controlled by multiple enhancers. For example, the *Drosophila yellow* and *white* genes considered later in this chapter are controlled by three enhancers—three clusters of contiguous binding sites for activators.

Enhancers that interact with many activators allow for very fine control over the expression of genes. Different combinations of activators produce different levels of expression of a given gene in different cells. In fact, the presence or absence of various enhancer elements near a gene reminds one of a binary code, where the presence is an "on" switch, and the absence is an "off" switch. Of course, the activators also have to be present to throw the switches. It may not be a simple additive arrangement, however, since multiple enhancer elements are known to act cooperatively.

Another metaphor that works well in describing the actions of multiple activators on multiple enhancer elements is a **combinatorial code.** The concentrations of all the activators in any given cell at a given time constitute the code. A gene can read the code if it has a battery of enhancer elements, each responsive to one or more of the activators. The result is an appropriate level of expression of the gene.

Eric Davidson and colleagues provided a beautiful example of multiple enhancer elements in the *Endo 16* gene of a sea urchin. This gene is active in the early embryo's vegetal plate—a group of cells that produces the endodermal tissues, including the gut. Davidson and colleagues began by testing DNA in the *Endo 16* 5′-flanking region for the ability to bind nuclear proteins. They found dozens of such regions, arranged into six modules, as illustrated in Figure 12.25.

How do we know that all these modules that bind nuclear proteins are actually involved in gene activation? Chiou-Hwa-Yuh and Davidson tested them by linking them alone and in combinations to the *cat* reporter gene (Chapter 5), reintroducing these constructs into sea urchin eggs, and observing the patterns of expression of the reporter gene in the resulting developing embryo. They found that the reporter gene was switched on in different parts of the embryo and at different times, depending on the exact combination of modules attached. Thus, the modules were responding to activators that were distributed nonuniformly in the developing embryo.

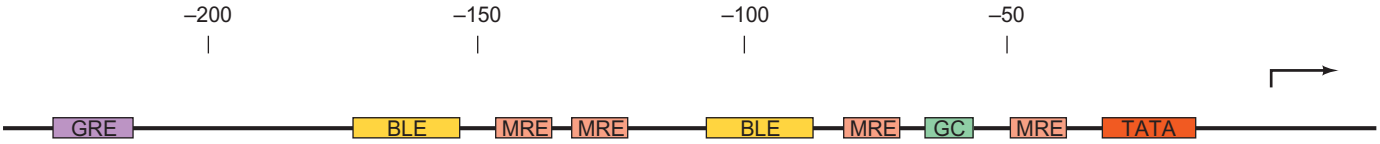


Figure 12.24 Control region of the human metallothionine gene. Upstream of the transcription start site at position +1 we find, in 3′–5′ order: the TATA box; a metal response element (MRE) that allows the gene to be stimulated in response to heavy metals; a GC box that responds to the activator Sp1; another MRE; a basal level enhancer (BLE) that responds to the activator AP-1; two more MREs; another BLE; and a glucocorticoid response element (GRE) that allows the gene to be stimulated by an activator composed of a glucocorticoid hormone and its nuclear receptor.

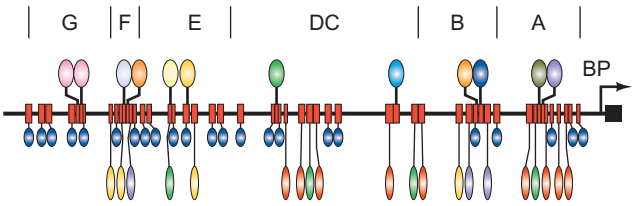


Figure 12.25 Modular arrangement of enhancers at the sea urchin *Endo 16* gene. The large colored ovals represent activators, and the small blue ovals represent architectural transcription factors, bound to enhancer elements (red boxes). The enhancers are arranged in clusters, or modules, as indicated by the regions labeled G, F, E, DC, B, and A. Long vertical lines denote restriction sites that define the modules. BP stands for "basal promoter." (*Source:* Adapted from Romano, L.A. and G.A. Wray, Conversation of *Endo 16* expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation, *Development 130 (17): 4189, 2003*.)

Figure 12.26 Control region of the human T-cell receptor α-chain (TCRα) gene. Within 112 bp upstream of the start of transcription lie three enhancer elements, which bind Ets-1, LEF-1, and CREB. These three enhancers are identified here by the transcription factors they bind, not by their own names.

Although all the elements may be able to function independently in vitro, the situation is more organized in vivo. Module A appears to be the only one that interacts directly with the basal transcription apparatus; all the other modules work through A. Some of the upstream modules (B and G) act synergistically through A to stimulate *Endo 16* transcription in endoderm cells. The other modules (DC, E, and F) act synergistically through A to block *Endo 16* transcription in nonendoderm cells (modules E and F play this role in ectoderm cells, and module DC plays this role in skeletogenic mesenchyme cells).

SUMMARY Complex enhancers enable a gene to respond differently to different combinations of activators. This arrangement gives cells exquisitely fine control over their genes in different tissues, or at different times in a developing organism.

Architectural Transcription Factors

The looping mechanism we have discussed for bringing together activators and general transcription factors is quite feasible for proteins bound to DNA elements that are separated by at least a few hundred base pairs because DNA is flexible enough to allow such bending. On the other hand, many enhancers are located much closer to the promoters they control, and that presents a problem: DNA looping over such short distances will not occur spontaneously, because short DNAs behave more like rigid rods than like flexible strings.

How then do activators and general transcription factors bound close together on a stretch of DNA interact to stimulate transcription? They can still approach each other if something else intervenes to bend the DNA more than the DNA itself would normally permit. We now have several examples of **architectural transcription factors** whose sole (or main) purpose seems to be to change the shape of a DNA control region so that other proteins can interact successfully to stimulate transcription. Rudolf Grosschedl and his colleagues provided the first example of a eukaryotic architectural transcription factor. They used the human T-cell receptor α-chain (TCRα) gene control region, which contains three enhancers, binding sites for the activators Ets-1, LEF-1, and CREB within just 112 bp of the transcription start site (Figure 12.26).

LEF-1 is the lymphoid enhancer-binding factor, which binds to the middle enhancer pictured in Figure 12.26 and helps activate the TCRα gene. However, previous work by Grosschedl and others had shown that LEF-1 by itself cannot activate TCRα gene transcription. So what is its role? Grosschedl and coworkers established that it acts by binding primarily to the minor groove of the enhancer and bending the DNA by 130 degrees.

These workers demonstrated minor groove binding by two methods. First, they showed that methylating six enhancer adenines on N3 (in the minor groove) interfered with enhancer function. Then they substituted these six A–T pairs with I–C pairs, which look the same in the minor groove, but not the major groove, and found no loss of enhancer activity. This is the same strategy Stark and Hawley used to demonstrate that TBP binds to the minor groove of the TATA box (Chapter 11).

Next, using the same electrophoretic assay Wu and Crothers used to show that CAP bends *lac* operon DNA (Chapter 7), Grosschedl and coworkers showed that LEF-1 bends DNA. They placed the LEF-1 binding site at different positions on linear DNA fragments, bound LEF-1, and measured the electrophoretic mobilities. The mobility was greatly retarded when the binding site was in the middle of the fragment, suggesting significant bending.

They also showed that the DNA bending is due to a so-called **HMG domain** on LEF-1. **HMG proteins** are small nuclear proteins that have a high electrophoretic mobility (hence, high mobility group, or HMG). To show the importance of the HMG domain of LEF-1, these workers prepared a purified peptide containing just the HMG domain and showed that it caused the same degree of bending (130 degrees) as the full-length protein. Extrapolation of the mobility curve to the point of maximum mobility (where the bend-inducing element should be right at the end of the

DNA fragment) indicated that the bend occurs at the LEF-1 binding site. Because LEF-1 does not enhance transcription by itself, it seems likely that it acts indirectly by bending the DNA. This presumably allows the other activators to contact the basal transcription machinery at the promoter and thereby enhance transcription.

SUMMARY The activator LEF-1 binds to the minor groove of its DNA target through its HMG domain and induces strong bending in the DNA. LEF-1, an architectural transcription factor, does not enhance transcription by itself, but the bending it induces probably helps other activators bind and interact with other activators and the general transcription factors to stimulate transcription.

Enhanceosomes

We have discussed several examples of enhancers, ranging from modular and spread out (the sea urchin *Endo 16* enhancer) to compact (the TCRα enhancer). We saw that transcription of the *Endo 16* gene responds differently to different combinations of activators, which also means that the *Endo 16* gene can be activated by subsets of activators. But not all enhancers work that way. Tom Maniatis and colleagues have studied an enhancer at the other end of the continuum of enhancer size and complexity: the human interferon-β (IFN-β) enhancer, which responds to viral infection. This enhancer contains binding sites for only eight polypeptides: two from the heterodimer ATF-2/cJun; four from two copies each of the interferon response factors IRF-3 and IRF-7; and two from the heterodimer **nuclear factor kappa B (NF6B),** whose two subunits are p50 and RelA. These proteins interact with proteins at the promoter through a coactivator known as CREB-binding protein (CBP), or its closely related cousin, p300.

In contrast to the *Endo 16* enhancer, the IFN-β enhancer works only when all of its activators are present at the same time in a cell. This is important because all of these activators activate many genes and are present in a wide variety of cells. Nevertheless, the IFN-β gene is strongly activated only when it is needed: when a cell is under attack by a virus. The requirement for all the activators at once explains this paradox, because all the activators are present together essentially only when cells are virus-infected.

Another protein that plays an important role in IFN-β activation is another member of the HMG family: **HMGA1a.** Unlike LEF-1, proteins of the HMGA1a type do not bend DNA. Instead they modulate the natural bending of A-T rich DNA regions. HMGA1a is essential for activation of the IFN-β gene, and its role is to ensure cooperative binding of the other activators to the enhancer.

The fact that the IFN-β enhancer binds several proteins cooperatively, and requires another protein that can modulate DNA bending, gave rise to the concept of the **enhanceosome,** a collection of proteins bound to an enhancer, all required for the complex to adopt a specific shape that can activate transcription efficiently. The original enhanceosome concept assumed that the DNA in an enhanceosome would be significantly bent, and that HMG proteins would play a role in such bending. However, we now know that HMGA1a does not bend DNA and, as we will soon see, it is not even part of the IFN-β enhanceosome, so the assumption of an enhanceosome with a strongly bent DNA rested on shaky ground.

Indeed, in 2007 Maniatis and colleagues assembled the crystal structure of the IFN-β enhanceosome (Figure 12.27) from two parts: The DNA-binding domains of IRF-3, IRF-7, and NF6B from one-half of the enhanceosome and a previously determined structure for the other half. They found that the DNA within the enhanceosome is essentially straight, experiencing only a gentle undulation. The IFN-β

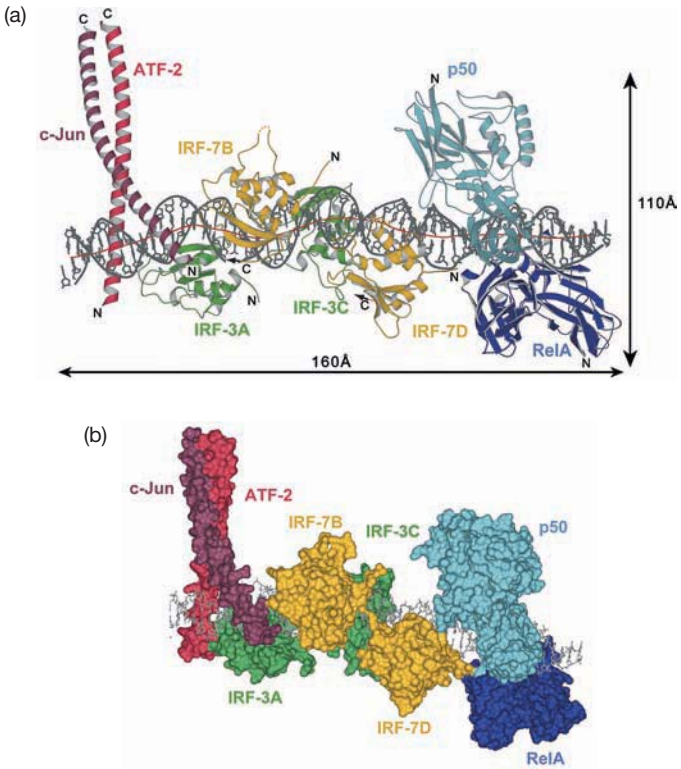


Figure 12.27 Model for the human IFN-β enhanceosome. **(a)** Ribbon diagram of the enhanceosome showing the gently undulating path of the DNA, whose local axis is traced by the dotted red line. The two IRF-3 molecules are designated -3A and -3C, and the two IRF-7 molecules are designated -7B and -7D. The overlapping binding sites for all the activators are shown on the DNA sequence below the diagram. **(b)** Molecular surface diagram of the enhanceosome in the same orientation as in panel (a). (*Source:* Reprinted from CELL, Vol. 129, Panne et al, An Atomic Model of the Interferon-β Enhanceosome, Issue 6, 15 June 2007, pages 1111–1123, © 2007, with permission from Elsevier.)

enhancer contains four binding sites for HMGA1a, but this protein is apparently not bound along with all the other activators. There is simply not room for it in the final enhancesome. But the crystal structure does emphasize the role of HMGA1a in cooperative binding of the other activators to the enhancer: It shows that, although the activators bind close together, they interact with each other to a remarkably small extent. Thus, HMGA1a presumably stimulates cooperativity by binding transiently to the DNA and other activators and helping them come together.

SUMMARY An enhanceosome is a nucleoprotein complex containing a collection of activators bound to an enhancer in such a way that stimulates transcription. The archetypical enhanceosome involves the IFN-β enhancer. Its structure involves eight polypeptides bound cooperatively to an essentially straight 55-bp stretch of DNA. HMGA1a is essential for this cooperative binding, but it is not part of the final enhanceosome.

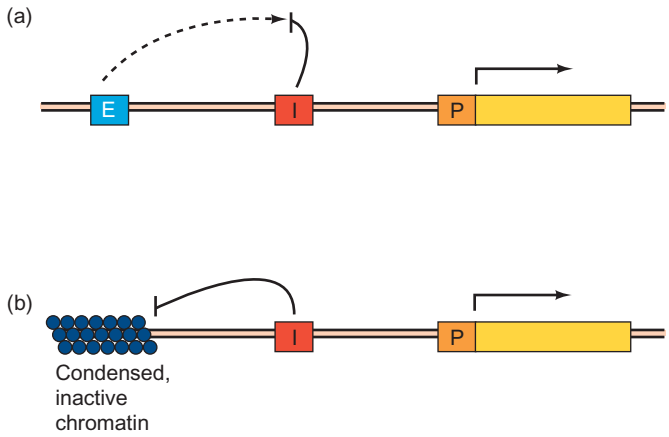


Figure 12.28 Insulator function. (a) Enhancer-blocking activity. The insulator between a promoter and an enhancer prevents the promoter from feeling the activating effect of the enhancer. **(b)** Barrier activity. The insulator between a promoter and condensed, repressive chromatin (induced by a silencer) prevents the promoter from feeling the repressive effect of the condensed chromatin (indeed, prevents the condensed chromatin from engulfing the promoter).

Insulators

We know that enhancers can act at a great distance from the promoters they activate. For example, the wing margin enhancer in the *Drosophila cut* locus is separated by 85 kb from the promoter. With a range that large, some enhancers will likely be close enough to other, unrelated genes to activate them as well. How does the cell prevent such inappropriate activation? Higher organisms, including at least *Drosophila* and mammals, use DNA elements called **insulators** to block activation of unrelated genes by nearby enhancers.

Gary Felsenfeld has defined an insulator as a "barrier to the influence of neighboring elements." An insulator that can protect a gene from activation by nearby enhancers is called an **enhancer blocking insulator.** On the other hand, an insulator that stops the encroachment of condensed chromatin into a target gene, thereby preventing gene silencing, is called a **barrier insulator.** Although many do, not all insulators have both blocking and barrier activities. Some are specialized for one activity or the other. The yeast elements that serve as barriers to the silencers at telomeres are prominent examples of insulators with only barrier activity.

How do insulators work? The details are not clear yet, but we do know that insulators define boundaries between DNA domains. Thus, an insulator abolishes activation if placed between an enhancer and a promoter. Similarly, an insulator abolishes repression if placed between a silencer and a silenced gene. It appears that the insulator creates a boundary between the domain of the gene and that of the enhancer (or silencer) so the gene can no longer feel the activating (or repressing) effects (Figure 12.28).

We also know that insulator function depends on protein binding. For example, certain *Drosophila* insulators contain the sequence GAGA and are known as **GAGA boxes.** These require the GAGA-binding protein **Trl** for insulator activity. Genetic experiments have shown that insulator activity can be abolished by mutations in either the GAGA box itself, or in the *trl* gene, which encodes Trl.

One can imagine many mechanisms for insulator function. We can easily eliminate one of these: a model in which the insulator induces a silenced, condensed chromatin domain upstream of the insulator. If that were the case, then a gene placed upstream of an insulator would always be silenced. But experiments with *Drosophila* have shown that such upstream genes are still potentially active and can be activated by their own enhancers.

Figure 12.29 illustrates two more models of insulator action. The first involves a signal that somehow moves progressively from the enhancer to the promoter, and the insulator blocks the progression of this signal. The second requires interaction between insulators on either side of an enhancer, which isolates the enhancer on a loop so it cannot interact with the promoter.

The first hypothesis is hard to reconcile with an experiment performed by J. Krebs and Dunaway similar in concept to the one by Dunaway and Dröge we discussed earlier in this chapter. In that earlier experiment (see Figure 12.20), Dunaway and Dröge placed a promoter and an enhancer on separate DNA circles linked in a catenane and showed that the enhancer still worked. In the later experiment, Krebs and Dunaway used the same catenane construct, but this time they surrounded either the enhancer or promoter with two *Drosophila* insulators: *scs* and *scs′*. They found that in both cases, the insulators blocked enhancer activity.

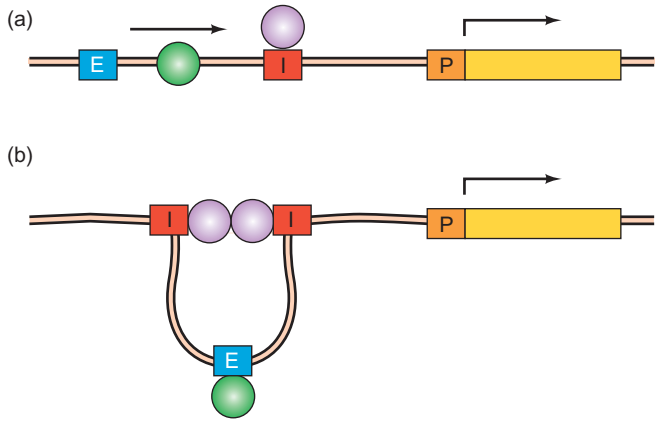


Figure 12.29 Two hypotheses for the mechanism of insulator activity. (a) Sliding model. An activator has bound to an enhancer and a stimulatory signal (green), perhaps the activator itself, is sliding along the DNA from the enhancer toward the promoter. But the insulator (red), perhaps with a protein or proteins attached, stands in the way and prevents the signal from reaching the promoter. **(b)** Looping model. Two insulators (red) flank an enhancer (blue). When proteins (purple) bind to these insulators, they interact with one another, isolating the enhancer on a loop so it cannot stimulate transcription from the nearby promoter (orange).

On the other hand, a single insulator in either circle had little effect on enhancement. Both experiments from Dunaway's group are incompatible with a signal propagating from the enhancer to the promoter unless the signal can jump from one DNA circle to another.

Arguments against the second hypothesis have been presented as well. Chief among them is the fact that some insulators work as single copies, so it is not apparent that there are two insulators flanking an enhancer. However, it is possible that the second insulator is present but not recognized in these experiments. It could attract novel proteins that can interact with the proteins that bind to the known insulator. Thus, the chromatin could be forced to loop in such a way as to prevent the enhancer from interacting with a promoter on one side, but not on the other.

Haini Cai and Ping Shen have performed experiments that support this hypothesis. When they placed a single copy of a known *Drosophila* insulator [su(Hw); (suppressor of *Hairy wing*)] between an enhancer and a promoter, they observed some insulator activity (a decrease in the effectiveness of the enhancer). However, when they placed two copies of the same insulator in the same place, they observed no insulator activity. Finally, when they placed single copies of the su(Hw) insulators on either side of the enhancer, they observed the most insulator activity of all. By the way, the Su(Hw) insulator is part of a retrotransposon (Chapter 23) known as *gypsy*. The insulator binds to a protein that is also known as Su(Hw).

Figure 12.30 illustrates Cai and Shen's interpretation of these results. Panel (a) shows what happened with the single insulator. It teamed up with an unknown insulator (I) somewhere upstream of the enhancer to block the action of the enhancer. Panel (b) shows what happened with an insulator on either side of the enhancer. Proteins bound to the insulators and caused the DNA to loop, isolating the enhancer in the loop in such a way that it could no longer interact with the promoter. In panel (c), the two adjacent insulators between enhancer and promoter bound proteins that interacted with each other, looping out the DNA in between, but

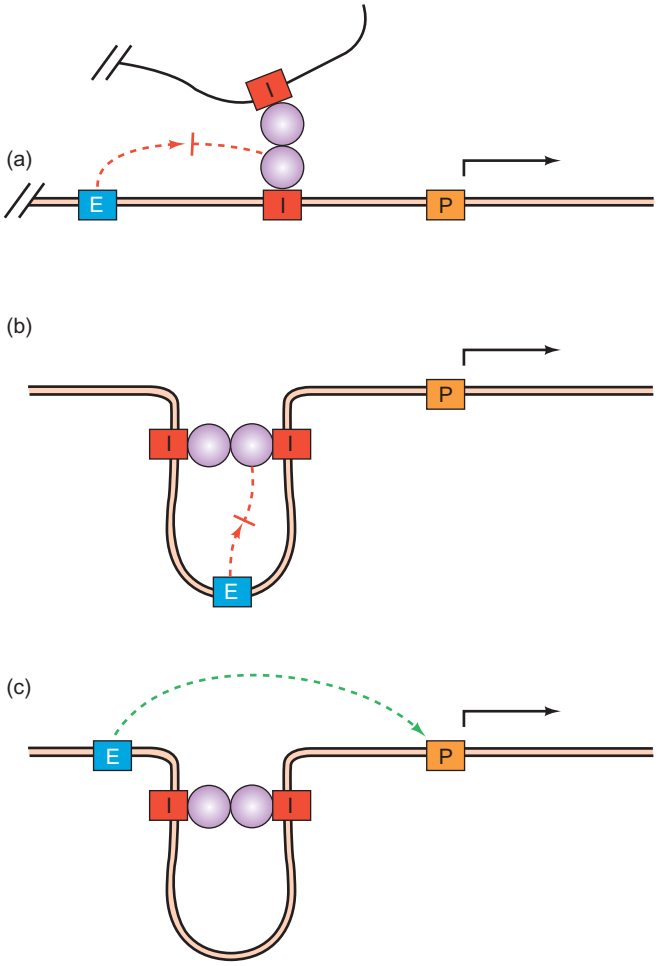


Figure 12.30 Model of multiple insulator action. (a) A single insulator (I, red) between an enhancer (E, blue) and a promoter (P, orange) binds to a protein(s) (purple) that interact with other protein(s), not necessarily of the same type, that are bound to another, remote insulator, also not necessarily of the same type. These protein–protein interactions isolate the enhancer from the promoter and block enhancement of transcription. **(b)** Two insulators flanking an enhancer bind to proteins that interact, looping the DNA and isolating the enhancer from the promoter. This prevents enhancement. **(c)** Two (or more) insulators between the enhancer and promoter bind to proteins that interact and loop out the DNA in between but do not isolate the enhancer from the promoter; in fact they bring the two elements closer together. Thus, the two insulators cancel each other out and do not block enhancement. The enhancer and promoter probably interact by DNA looping that is not illustrated here. (*Source:* Adapted from Cai, H.N. and P. Shen, Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* 291 [2001] p. 495, (4.))

did not interfere with enhancer activity. In fact, the looped DNA actually brought the enhancer closer to the promoter and presumably made the enhancer more effective.

At the same time in 2001, Vincenzo Pirrotta and coworkers reported work in which they performed the same kind of experiment with the su(Hw) insulator in single and multiple copies, but with different *Drosophila* promoters, and obtained the same results. Then they added a new wrinkle: two different genes in tandem, instead of just one, with three upstream enhancers, and one to three insulators in various positions. The two genes were *yellow* and *white,* which are responsible for dark body and wing color, and for red eye color, respectively, in adult flies. When the *yellow* gene is inactivated (or mutated) dark pigment fails to be made, and the body and wings are yellow instead of black. When the *white* gene is inactivated (or mutated), red eye pigment synthesis fails and the eyes of the fly are white.

Figure 12.31 illustrates the constructs Pirrotta and coworkers used, and the results they obtained. The first construct (EyeSYW) contained one copy of the insulator between the enhancers and the two genes. As Cai and Shen's model predicted, the insulator prevented activation of both genes by the enhancers. The second construct (EyeYSW) contained an insulator between the *yellow* and *white* genes. Again, predictably, the *yellow* gene was activated, but the *white* gene was not.

The third construct (EyeSYSW), in which two insulators flanked the *yellow* gene, is more interesting. This time, the *yellow* gene was not activated, but the *white* gene was. Again, Cai and Shen's model is compatible with these results: The two insulators flanking the *yellow* gene prevented its activation, but they constituted two insulators together between the enhancers and the *white* gene, so they cancelled each other and allowed activation of that gene. Thus, the interaction of the two insulators, while it cancelled their effect on the *white* gene, did not really inactivate them: They could still prevent inactivation of the *yellow* gene that lay between them. The fourth construct (EyeSYWS) contained two insulators flanking the *yellow* and *white* genes. Predictably, the insulators prevented activation of both genes.

Finally, the fifth construct (EyeSFSYSW) contained three insulators, two between the enhancers and the *yellow* gene, and one between the *yellow* and *white* genes. Because both genes were activated, we see that two *or more* copies of the insulator between an enhancer and a gene neutralizes the effect of the insulators. (There are two copies between the enhancers and the *yellow* gene, but three between the enhancers and the *white* gene.) We might have expected the two insulators upstream of the *yellow* gene to neutralize each other and allow activation of the *yellow* gene, but the single remaining insulator between the *yellow* and *white* genes might have been expected to block activation of the *white* gene. Instead, none of the three insulators had any effect, and both genes were activated. This experiment therefore revealed that the inactivation of two tandem insulators is not due to a simple, exclusive interaction between the two. Somehow, proteins bound to all three

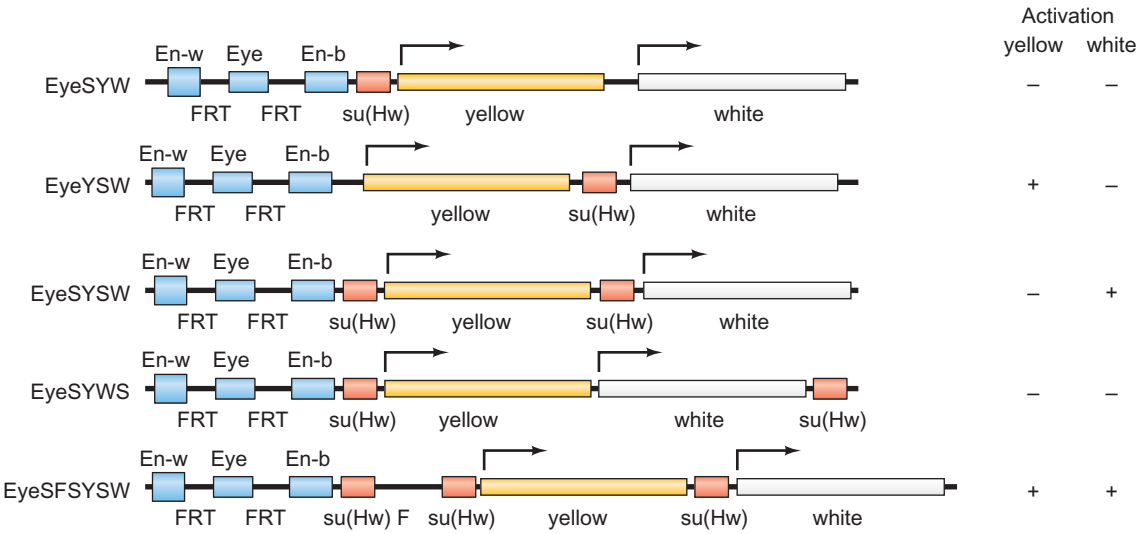


Figure 12.31 Effects of insulators on two tandem *Drosophila* genes. The structures of the constructs are given on the left, with the results (activation [+] or no activation [–] of the *yellow* and *white* genes) on the right. The names of the constructs all begin with Eye, which stands for the eye-specific enhancer found in the cluster of three enhancers (blue) upstream of both the *yellow* and *white* genes. The S, Y, and W in the names stand for the insulator [su(Hw), red], the *yellow* gene, and the *white* gene, respectively. The F in the last construct stands for a spacer fragment. The positions of the letters in the construct names indicate the positions of the corresponding elements in the constructs. Pirrotta and coworkers placed each construct into *Drosophila* embryos and observed the effects on body and wing color (*yellow* gene activity) and on eye color (*white* gene activity). (*Source:* Adapted from Muravyova, E., A. Golovnin, E. Gracheva, A. Parshikov, T. Belenkaya, V. Pirotta, and P. Georgiev, Loss of insulator activity by paired Su(Hw) Chromatin Insultators. *Science* 291 [2001] p. 497, f. 2.)

insulators appear to interact in such a way as to permit the enhancers upstream to do their job.

All of these results on enhancement and insulator action are easiest to explain on the basis of DNA looping, as illustrated in Figure 12.30. But looping is not the only possible explanation. Experimental evidence to date cannot rule out some kind of tracking mechanism (see Figure 12.18d) to explain enhancement. And proteins bound to the enhancer and tracking toward the promoter would be readily blocked by placing a single insulator between the enhancer and the promoter. How then can we explain the canceling effect of two or more insulators between the enhancer and the promoter? One way is to invoke **insulator bodies,** which are conglomerations of two or more insulators and their binding proteins that have been detected at the periphery of the nucleus. The formation of insulator bodies is thought to play a critical role in insulator activity, but we have no accepted hypothesis for how the insulator bodies play this role. In the absence of such a hypothesis, we cannot rule out the possibility that two or more insulators (lying between an enhancer and a promoter) and their binding proteins interact with each other in such a way as to prevent the association of the insulators with insulator bodies. And such interactions would thereby block insulator activity.

Another model for insulator activity, proposed by Pfeifer and colleagues, is that the insulator blocks association between enhancers and promoters by forming associations of its own with these chromosomal elements. Of course, it is not the DNA regions themselves, but the proteins bound to these DNA regions, that are interacting. As we learned earlier in this chapter, Pfeifer and colleagues showed that the *Igf2* enhancers and promoters are brought together by DNA looping when the gene is activated, but not when it is silenced. Furthermore, we learned that the maternal copy of the gene is silenced by imprinting (see Box 12.1), while the paternal copy remains active in fetal muscle and liver cells.

It was already known by 2007 that silencing of the maternal *Igf2* gene depended on the **imprinting control region** (***ICR,*** refer back to Figure 12.22a). Furthermore, the *ICR* silences the maternal gene by acting as an insulator that shields the maternal *Igf2* promoters from the stimulatory effects of the two nearby enhancers. The *ICR* insulator binds to **CTCF** (CCCTC-binding factor), which is a common insulator-binding protein that interacts with a variety of insulators found throughout vertebrate genomes. Pfeifer and colleagues, and others, had previously shown that removal of the *ICR* from the maternal chromosome allowed expression of the maternal copy of the *Igf2* gene. Then, Pfeifer and colleagues demonstrated (by the same kind of 3C and RFLP analysis shown in Figure 12.22) that removal of the *ICR* from the maternal chromosome also allowed the maternal enhancers to associate with the *Igf2* promoters. This bolstered the hypothesis that physical association between enhancers and promoters is essential for enhancer activity, and the *ICR* insulator acts by blocking that essential association.

But how does the *ICR* insulator block association between the *Igf2* enhancers and promoters? Pfeifer and colleagues proposed that CTCF bound to the insulator interacts with the enhancers and promoters, or proteins bound to both, and prevents their interaction with each other (Figure 12.32). To test this hypothesis, they performed 3C and RFLP analysis on maternal and paternal chromosomes, with and without the insulator, and showed that indeed the insulator interacts with both enhancers and promoters, but only on the maternal chromosome, in which *Igf2* transcription is silenced.

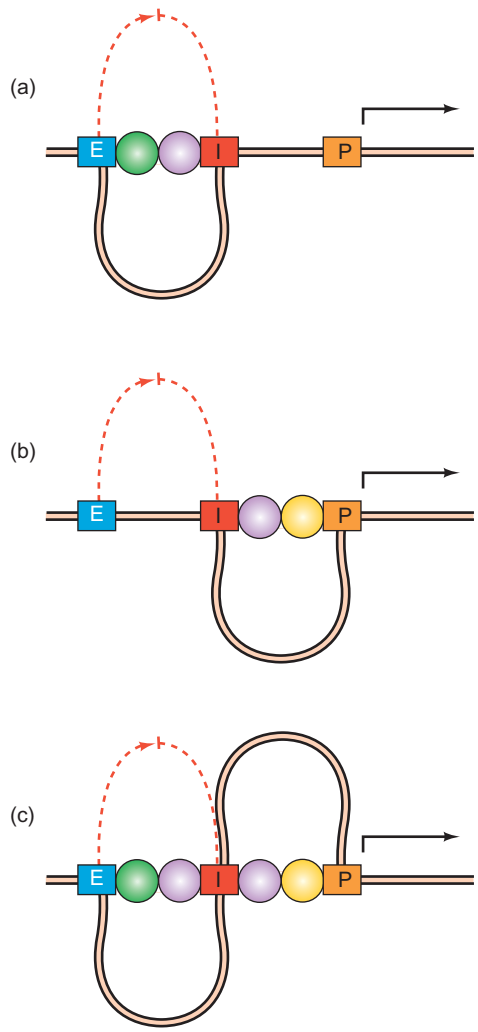


Figure 12.32 Model for insulator action by binding to enhancers and/or promoters. (a) The insulator binds to an enhancer (through proteins bound to both) and prevents its interaction with a promoter. **(b)** The insulator binds to a promoter (again through proteins) and prevents its interaction with an enhancer. **(c)** The insulator binds to both promoter and enhancer (through proteins) and prevents interaction between the promoter and enhancer.

Thus, in this system at least, insulator action appears to depend on the insulator's interacting with the enhancers and promoters in such a way that they cannot interact with each other. In some ways, this is an attractive hypothesis, but it has serious limitations as a general explanation for insulator action. First, insulators are position dependent. They block enhancer action only when placed between the enhancer and a promoter. In the present example, the *ICR* insulator blocks the enhancers from stimulating transcription from the *Igf2* promoters, but not from the *H19* promoter. It is not obvious why the position of the *ICR* insulator between the *Igf2* promoters and enhancers would cause it to interact only with those promoters, and not the *H19* promoter, which is much closer to the insulator. Second, insulators do not inactivate enhancers. While they block the action of an enhancer on one set of promoters (e.g., the *Igf2* promoters), they leave it free to stimulate transcription from another (e.g., the *H19* promoter). It is not clear how binding of the insulator to the *Igf2* enhancers and promoters would prevent their interaction with each other, and still allow them to interact productively with other chromosomal partners such as the *H19* promoter.

Finally, you may be wondering why the paternal copy of the *Igf2* gene is not affected by the insulator. The paternal *ICR* becomes methylated during and after spermiogenesis, so it cannot bind CTCF. Without the insulator-binding protein, the insulator cannot function, so the enhancers are allowed to stimulate transcription from the paternal *Igf2* promoters. Thus, methylation of the insulator is the functional equivalent of its removal.

Perhaps the best way to summarize our knowledge about the mechanism of insulator action is to acknowledge that there may not be a single mechanism. Some insulators may work one way, and others may have another mode of action.

SUMMARY Insulators are DNA elements that can shield genes from activation by enhancers (enhancer-blocking activity) or repression by silencers (barrier activity). Some insulators have both enhancer-blocking and barrier activities, but some have only one or the other. Insulators may do their job by working in pairs that bind proteins that can interact to form DNA loops. These loops would isolate enhancers and silencers so they can no longer stimulate or repress promoters. In this way, insulators may establish boundaries between DNA regions in a chromosome. Two or more insulators between an enhancer and a promoter cancel each other's effect, perhaps by binding proteins that interact with each other, thereby preventing the DNA looping that would isolate the enhancer from the promoter. Alternatively, the interaction between adjacent insulator-binding proteins could prevent the association of the insulators with insulator bodies, and this could block insulator activity. Insulators may also act as a barrier to a signal propagating along the chromosome from an enhancer or silencer. The nature of this signal is not defined, but it may be a sliding protein or a sliding (and growing) loop of chromatin. Finally, enhancer-blocking insulators may act by binding proteins that interact with proteins and/or DNA at enhancers and promoters, thereby preventing those enhancers and promoters from interacting with each other, which is essential for efficient transcription.

12.6 Regulation of Transcription Factors

Transcription factors regulate transcription both positively and negatively, but what regulates the regulators? We have already seen one example earlier in this chapter, and we will see several other examples in the last section of this chapter and in Chapter 13. They fall into the following categories:

- As we learned earlier in this chapter, binding between nuclear receptors (e.g., the glucocorticoid receptor) and their ligands (e.g., the glucocorticoids) can cause the receptors to dissociate from an inhibitory protein in the cytoplasm, translocate to the nucleus, and activate transcription.
- As we will see in Chapter 13, binding between nuclear receptors and their ligands can change the receptors from transcription repressors to activators.
- Phosphorylation of activators can allow them to interact with coactivators that in turn stimulate transcription.
- Ubiquitylation of transcription factors (attachment of the polypeptide ubiquitin to them) can mark them for destruction by proteolysis.
- Alternatively, ubiquitylation of transcription factors can stimulate their activity instead of marking them for destruction.
- Sumoylation of transcription factors (attachment of the polypeptide SUMO to them) can target them for incorporation into compartments of the nucleus where their activity cannot be expressed.
- Methylation of transcription factors can modulate their activity.
- Acetylation of transcription factors can modulate their activity.

Let us examine some of these regulation phenomena.

Coactivators

Some class II activators may be capable of recruiting the basal transcription complex all by themselves, possibly by contacting one or more general transcription factors or RNA polymerase. But many, if not most, cannot. Roger Kornberg and colleagues provided the first evidence that something else must be involved when they studied **activator interference,** or **squelching,** in 1989 and 1990. Squelching occurs when increasing the concentration of one activator inhibits the activity of another activator in an in vitro transcription experiment, presumably by competing for a scarce factor required by both activators. A reasonable candidate for such a limiting factor would be a general transcription factor, but Kornberg and coworkers discovered that adding very large quantities of the general transcription factors did not relieve squelching. This finding suggested that some other factor must be required by both activators.

What was this other factor? In 1990, Kornberg and colleagues partially purified a yeast protein that could relieve squelching. Then, in 1991, they purified this factor further and demonstrated directly that it had coactivator activity. That is, it could stimulate activated transcription, but not basal transcription in vitro. They called it **Mediator** because it appeared to mediate the effect of an activator. (We have already encountered Mediator in Chapter 11 in the context of the polymerase II holoenzyme.)

Kornberg and colleagues' assay for transcription used a G-less cassette (Chapter 5) driven by the yeast *CYC1* promoter and a GAL4-binding site. They added increasing concentrations of Mediator in the absence and presence of the activator GAL4-VP16, a chimeric activator with the DNA-binding domain of GAL4 and the transcription-activating domain of VP16. Figure 12.33 shows the results: Mediator had no effect on transcription in the absence of the activator (lanes 3–6), but it greatly stimulated transcription in the presence of the activator (lanes 7–10). A similar experiment with the yeast activator GCN4 yielded comparable results, showing that Mediator could cooperate with more than one activator having an acidic activation domain.

Mediator-like complexes have also been purified from higher eukaryotes, including humans. One such complex has been purified independently by two different groups and is therefore called by two different names: **SRB and MED-containing cofactor (SMCC),** and **thyroid-hormone-receptor-associated protein (TRAP).** SMCC/TRAP is the most complex of the known Mediator-like complexes in mammals, but there are others that seem to be structurally and functionally related to Mediator. One of these is CRSP, which we will discuss later in this section.

Further work has shown that Mediator and its homologs are ubiquitous participants at active class II promoters. Indeed, they are so widespread that they can be considered general transcription factors, rather than true coactivators. A typical **coactivator** is a protein that has no activator function of its own, but collaborates with one or more activators to stimulate the expression of a set of genes.

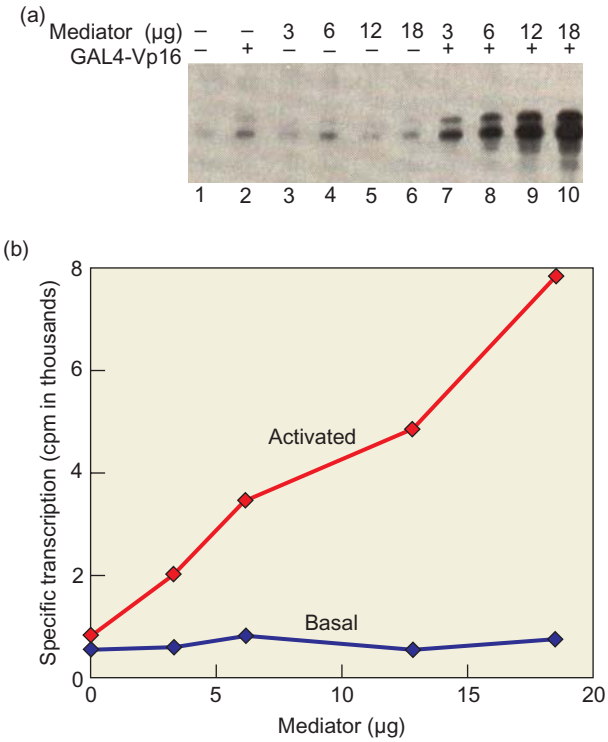


Figure 12.33 Discovery of Mediator. Kornberg and colleagues placed the yeast *CYC1* promoter downstream of a GAL4-binding site and upstream of a G-less cassette, so transcription of the G-less cassette depended on both the *CYC1* promoter and GAL4. Then they transcribed this construct in vitro in the absence of GTP and in the presence of the amounts of Mediator shown at the top of panel (a), and in the absence (−) or presence (+) of the activator GAL4-VP16 as indicated at the top of panel (a). They included a labeled nucleotide to label the products of the in vitro transcription reactions and electrophoresed the labeled RNAs. **(a)** Phosphorimager scan of the electropherogram. **(b)** Graphical presentation of the results in panel (a). Note that Mediator greatly stimulates transcription in the presence of the activator, but has no effect on unactivated (basal) transcription. (*Source:* Flanagan, P.M., R.J. Kelleher, 3rd, M.H. Sayre, H. Tschochner, and R.D. Kornberg, A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350 (4 Apr 1991) f. 2, p. 437. Copyright © Macmillan Magazines Ltd.)

For example, in Chapter 7 we learned that cyclic-AMP (cAMP) stimulates transcription of bacterial operons by binding to an activator (CAP) and causing it to bind to activator target sites in the operon control regions. Cyclic-AMP also participates in transcription activation in eukaryotes, but it does so in a less direct way, through a series of steps called a **signal transduction pathway.** When the level of cAMP rises in a eukaryotic cell, it stimulates the activity of **protein kinase A (PKA)** and causes this enzyme to move into the cell nucleus. In the nucleus, PKA phosphorylates an activator called the **cAMP response element-binding protein (CREB),** which binds to the **cAMP response element (CRE)** and activates associated genes.

Because phosphorylation of CREB is necessary for activation of transcription, one would expect this phosphorylation

to cause CREB to move into the nucleus or to bind more strongly to CREs, but neither of these things actually seems to happen—CREB localizes to the nucleus and binds to CREs very well even without being phosphorylated. How, then, does phosphorylation of CREB cause activation? The key to the answer appeared in 1993 with the discovery of the **CREB-binding protein (CBP).** CBP binds to CREB much more avidly after CREB has been phosphorylated by protein kinase A. Then, CBP can contact and recruit elements of the basal transcription apparatus, or it could recruit the holoenzyme as a unit. By coupling CREB to the transcription apparatus, CBP acts as a coactivator (Figure 12.34).

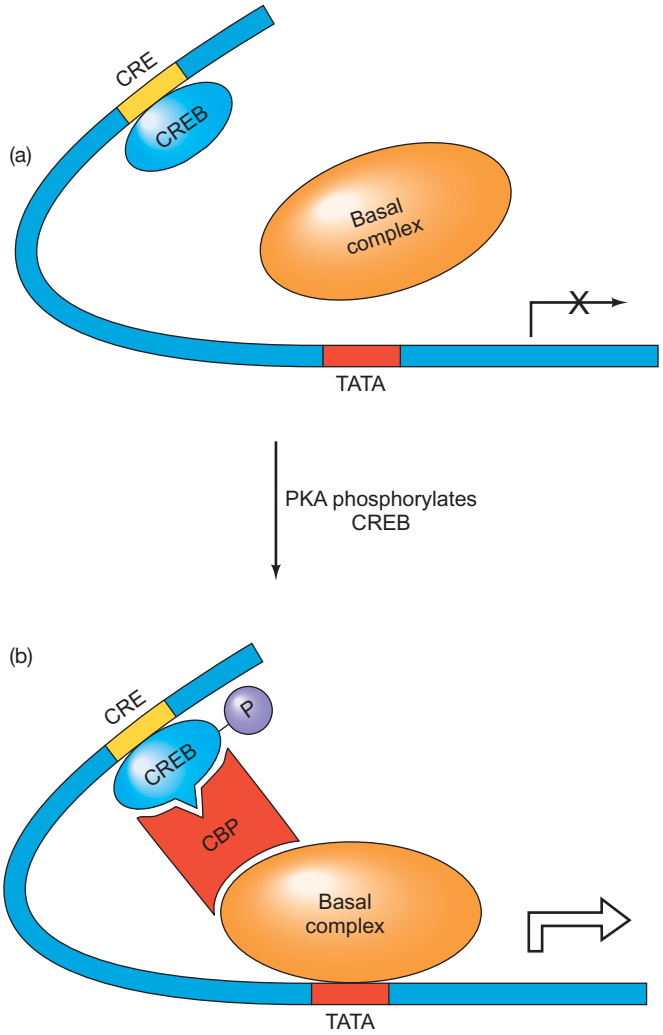


Figure 12.34 A model for activation of a CRE-linked gene. **(a)** Unphosphorylated CREB (turquoise) is bound to CRE, but the basal complex (RNA polymerase plus general transcription factors, orange) is not bound to the promoter in significant quantity and may not even have assembled yet. Thus, the gene is not activated. **(b)** PKA has phosphorylated CREB, which causes CREB to associate with CBP (red). CBP, in turn, associates with at least one component of the basal transcription complex, recruiting it to the promoter. Now transcription is activated.

Since 1993 when CBP was discovered, many coactivators, have been identified. In 1999, Tjian and colleagues isolated a coactivator required for activation of transcription in vitro by the transcription factor Sp1. When they purified this coactivator, which they called **cofactor required for Sp1 activation (CRSP),** they discovered that it had nine putative subunits. They separated these subunits by SDS-PAGE, transferred them to a nitrocellulose membrane, then cleaved each polypeptide with a protease to generate peptides that could be sequenced. The sequences revealed that some of the subunits of CRSP are unique, but many of them are identical, or at least homologous, to other known coactivators—subunits of the yeast Mediator, for example. Thus, different coactivators seem to be assembled by "mixing and matching" subunits from a variety of other coactivators. Mediator and CRSP also seem to share a mode of action in common. Both contact the CTD of RNA polymerase II. That interaction may explain how these coactivators help recruit the basal transcription complex.

The coactivator role of CBP is not limited to cAMP-responsive genes. It also serves as a coactivator in genes responsive to the nuclear receptors. This helps to explain why no one could detect direct interaction between the transcription-activation domains of the nuclear receptors and any of the general transcription factors. Part of the reason is that the nuclear receptors do not contact the basal transcription apparatus directly. Instead, CBP, or its homologue, **p300,** acts as a coactivator, helping to bring together the nuclear receptors and the basal transcription apparatus. But CBP does not perform this task alone. It collaborates with another family of coactivators called the **steroid receptor coactivator (SRC) family.** This group of proteins is also sometimes called the p160 family because of their molecular masses of 160 kD. The SRC family includes three groups of homologous proteins, **SRC-1, SRC-2,** and **SRC-3,** which interact with liganded (but not ligand-free) nuclear receptors. This interaction occurs between the nuclear receptor's activation domain and a so-called LXXLL box (where L stands for leucine and X stands for any amino acid) in the middle of the SRC protein chain. The SRC proteins also bind to CBP and can therefore help the nuclear receptors recruit CBP, which in turn recruits the basal transcription apparatus. The first SRC family member to be discovered was SRC-1 (Figure 12.35). It interacts with the ligand-bound forms of: progestin receptor; estrogen receptor; and thyroid hormone receptor. Not only does it bridge between nuclear receptors and CBP, it recruits a protein called **coactivator-associated arginine methyltransferase (CARM1),** which methylates proteins in the vicinity of the promoter, activating transcription. We will examine the role of CARM1 later in this section.

Still another important class of activators use CBP as a coactivator. A variety of growth factors and cellular stresses initiate a cascade of events (another signal transduction pathway) that results in the phosphorylation and activation

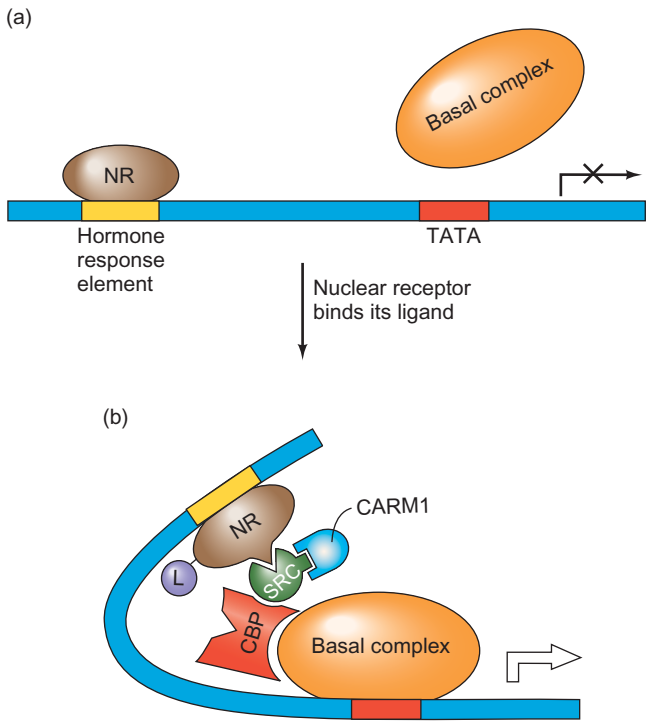


Figure 12.35 Models for activation of a nuclear receptor-activated gene. (a) A nuclear receptor (without its ligand) is bound to its hormone response element, but it cannot contact the basal transcription complex, so the linked gene is not activated. Depending on the type, the nuclear receptor could also be dissociated from its DNA target in the absence of its ligand. The nuclear receptor bound to its DNA target without its ligand may also actively inhibit transcription. **(b)** The nuclear receptor has bound to its ligand (purple) and is now able to interact with SRC (green), which in turn binds to CBP, which binds to at least one component of the basal transcription apparatus, recruiting it to the promoter and activating transcription. SRC also binds to CARM1 (torquoise), which methylates proteins near the promoter, further simulating transcription.

of a protein kinase called **mitogen-activated protein kinase (MAPK).** The activated MAPK enters the nucleus and phosphorylates activators such as Sap-1a and the Jun monomers in AP-1. These activators then use CBP to mediate activation of their target genes, which finally stimulate cell division.

Besides recruiting the basal transcription apparatus to the promoter, CBP plays another role in gene activation. CBP has a powerful histone acetyltransferase activity, which adds acetyl groups to histones. As we will see in Chapter 13, histones are general repressors of gene activity. Moreover, acetylation of histones causes them to loosen their grip on DNA and relax their repression of transcription. Thus, the association between activators and CBP at an enhancer brings the histone acetyltransferase to the enhancer, where it can acetylate histones and activate the nearby gene. We will discuss this phenomenon in greater detail in Chapter 13.

We have seen that CBP and p300 can serve as a coactivator for a variety of activators, including CREB and nuclear receptors. This means that the CREB and nuclear receptor pathways could potentially compete with each other for activation of different genes through the same coactivator. Ronald Evans and colleagues discovered that one way cells limit that competition is through methylation of CBP or p300. To simplify our discussion of this mechanism, we will refer to these proteins as CBP/p300.

Nuclear receptors attract not only CBP/p300, but several other proteins as well. One of these others is CARM1. The CARM1 activity methylates arginines on histones after they have been acetylated by CBP/p300, (Chapter 13) and this methylation has a transcription-activating effect. But CARM1 also methylates an arginine on CBP/p300 itself. The target arginine on CBP/p300 is in the so-called *KIX domain,* which is necessary for recruitment of CREB, but has no effect on the nuclear receptor-CBP/p300 interaction. Thus, CARM1 serves as a transcriptional switch. By blocking interaction between CBP/300 and CREB, CARM1 represses CREB-responsive genes, but CARM1 activates nuclear receptor-responsive genes by methylating histones in the vicinity.

SUMMARY Several different activators, including CREB, the nuclear receptors, and AP-1, do not activate transcription by contacting the basal transcription apparatus directly. Instead, they contact a coactivator called CBP (or its homolog p300), which in turn contacts the basal transcription apparatus and recruits it to promoters. CBP/p300 bound to nuclear receptor-response elements can also recruit CARM1, which methylates an arginine on CBP/p300 required to interact with CREB. This prevents activation of CREB-responsive genes.

Activator Ubiquitylation

Sometimes genes are inactivated by destruction of the activators that have been stimulating their activity. For example, transcription factors in the **LIM homeodomain (LIM-HD)** family associate with corepressors and coactivators. The coactivators are called **CLIM,** for "cofactor of LIM," among other names, and the corepressors are called **RLIM,** for "RING finger LIM domain-binding protein."

CLIM proteins are able to compete with RLIM proteins for binding to LIM-HD activators, so how do the RLIM proteins ever get the upper hand and repress LIM-HD-activated genes? The secret appears to lie in the ability of RLIM proteins to cause the destruction of LIM-HD-bound CLIM proteins, and thereby replace them. RLIM proteins set CLIM proteins up for destruction by binding to them and attaching several copies of a small protein called **ubiquitin** to lysine residues of the protein, creating what we call a **ubiquitylated protein.** Once the chain of ubiquitin molecules becomes long enough, it targets the ubiquitylated protein to a cytoplasmic structure called the **proteasome.** The proteasome is a collection of proteins with a combined

sedimentation coefficient of 26S. It includes proteases that degrade any ubiquitylated protein brought to it.

The normal function of the ubiquitin-linked proteasome appears to be quality control. It is estimated that about 20% of cellular proteins are made incorrectly because of mistakes in transcription or translation. These aberrant proteins are potentially damaging to the cell, so they are tagged with ubiquitin and sent to the proteasome for degradation before they can cause any trouble. Other proteins that are made correctly can become denatured by stresses such as oxidation or heat. The cell has **chaperone proteins** that can unfold and then allow such denatured proteins to refold correctly. But sometimes the denaturation is so extensive that proper refolding is impossible. In such cases, the denatured proteins would be ubiquitylated and then destroyed by the proteasome.

It may seem surprising that ubiquitylation can also affect activators without causing their destruction. One example comes from the *MET* genes of yeast, which are required to produce the sulfur-containing amino acids methionine and cysteine. These genes are controlled by the concentration of the methyl donor *S*-adenosylmethionine, known as SAM or AdoMet (Chapter 15). When the concentration of SAM is low, the *MET* genes are stimulated by the activator Met4. However, when the concentration of SAM rises, Met4 is inactivated by a process that involves ubiquitylation. This seems to imply that Met4 is ubiquitylated and then destroyed by the proteasome. However, things are not that simple.

It is true that Met4 degradation can play a role in its inactivation, but under certain conditions (rich medium supplemented with methionine), Met4 remains stable despite being ubiquitylated. However, even though it is stable, ubiquitylated Met4 loses its ability to activate the *MET* genes. It can no longer bind properly to these genes, even though it is still able to bind and activate another class of genes called the *SAM* genes. Thus, ubiquitylation of Met4 can inactivate it directly, without causing its destruction. And this inactivation is selective. It affects the ability of Met4 to activate some genes, but not others.

Several studies have indicated that very strong transcription factors tend to be regulated by ubiquitylation and subsequent destruction by the proteasome. This allows a cell some flexibility in controlling gene expression because it provides a mechanism for quickly shutting off strong expression of genes driven by powerful activators. But, again, the picture is not quite as simple as just protein degradation. Some of these activators are actually activated by monoubiquitylation (tagging the protein with a single copy of ubiquitin). But polyubiquitylation of the same activator can mark it for destruction.

Recently, evidence has accumulated for another kind of involvement of the proteasome in transcription regulation. Proteins belonging to the 19S regulatory particle of the proteasome have been discovered in complexes with transcription factors at active promoters. Moreover, the **19S particle** can strongly stimulate transcription elongation in vitro. Also, a subset of proteins from the 19S particle can be recruited to promoters by the activator GAL4. These proteins include ATPases that are necessary for unfolding proteins prior to their degradation but not proteins involved in proteolysis itself. Thus, the activation effect of the 19S particle proteins appears to be independent of proteolysis. Joan Conaway and colleagues speculated that the proteasomal proteins stimulate transcription by at least partially unfolding transcription factors so that they can be remodeled in such a way that stimulates transcription initiation, or elongation, or both.

SUMMARY RLIM proteins, which are LIM-HD corepressors, can bind to LIM-HD coactivators such as CLIM proteins and ubiquitylate them. This marks the coactivators for destruction by the 26S proteasome and allows the RLIM corepressors to take their place. Ubiquitylation (especially monoubiquitylation) of some activators can have an activating effect, but polyubiquitylation marks these same proteins for destruction. Proteins from the 19S regulatory particle of the proteasome can stimulate transcription, perhaps by remodeling and thereby activating transcription factors.

Activator Sumoylation

Sumoylation is the addition of one or more copies of the 101-amino-acid polypeptide **SUMO** (**small ubiquitin-related modifier**) to lysine residues on a protein. This process is accomplished by a mechanism very similar to the one used in ubiquitylation, but the results are quite different. Instead of being destroyed, sumoylated activators appear to be targeted to a specific nuclear compartment that keeps them stable, but unable to reach their target genes.

For example, certain activators, including one called PML, for "promyelocytic leukemia," are normally sumoylated and sequestered in nuclear bodies called PML oncogenic domains (PODs). In promyelocytic leukemia cells, the PODs are disrupted, and the released transcription factors, including PML, presumably reach and activate their target genes, and this activation contributes to the leukemic state.

Another example involves the Wnt signal transduction pathway, which ends when an activator called β-catenin enters the nucleus and teams up with LEF-1, an architectural transcription factor that we discussed earlier in this chapter, to activate transcription of certain genes. LEF-1 is subject to sumoylation, which causes it to be sequestered in nuclear bodies. Without LEF-1, β-catenin cannot activate its target genes, and Wnt signaling is blocked. And, as we have already learned, LEF-1 is involved in activating other genes, such as the TCR-α gene, independent of Wnt signaling, and those activations are also blocked by LEF-1 sumoylation.

Another consideration is that LEF-1 can also partner with repressors, such as Groucho, and this repression is presumably also blocked by sumoylation of LEF-1.

SUMMARY Some activators can be sumoylated (coupled to a small protein called SUMO), which causes them to be sequestered in nuclear bodies, where they cannot carry out their transcription activation function.

Activator Acetylation

In Chapter 13 we will learn that basic proteins called histones associate with DNA and repress transcription. It has been known for a long time that these histones can be acetylated on lysine residues by enzymes called histone acetyltransferases (HATs), which decreases the histones' repressive activity. Recently, investigators have shown that HATs can also acetylate nonhistone activators and repressors, and this can have either positive or negative effects on the acetylated protein's activities.

The tumor suppressor protein p53 is an example of an activator whose acetylation stimulates its activity. The coactivator p300 has HAT activity that can acetylate p53. When this happens, the activity of p53 increases, resulting in stronger stimulation of transcription of this activator's target genes.

The HAT activity of p300 can also acetylate the repressor BCL6, and this acetylation inactivates the repressor. Thus, the result in this case, as with the activation of an activator, is stimulation of transcription.

SUMMARY Nonhistone activators and repressors can be acetylated by HATs, and this acetylation can have either positive or negative effects.

Signal Transduction Pathways

The phosphorylations of CREB, Jun, and β-catenin, mentioned in the preceding section, are all the results of signal transduction pathways. So signal transduction pathways play a major role in the control of transcription. Let us explore the concept of signal transduction further and examine some examples. Cells are surrounded by a semipermeable membrane that keeps the cell contents from escaping and provides some protection from noxious substances in the cell's environment. This barrier between the interior of a cell and its environment means that mechanisms had to evolve to allow cells to sense the conditions in their surroundings and to respond accordingly. Signal transduction pathways provide these mechanisms. Because the responses a cell makes to its environment usually require changes in gene expression, signal transduction pathways usually end with activation of a transcription factor that activates a gene or set of genes.

Figure 12.36 outlines three signal transduction pathways: the protein kinase A pathway; the Ras–Raf pathway;

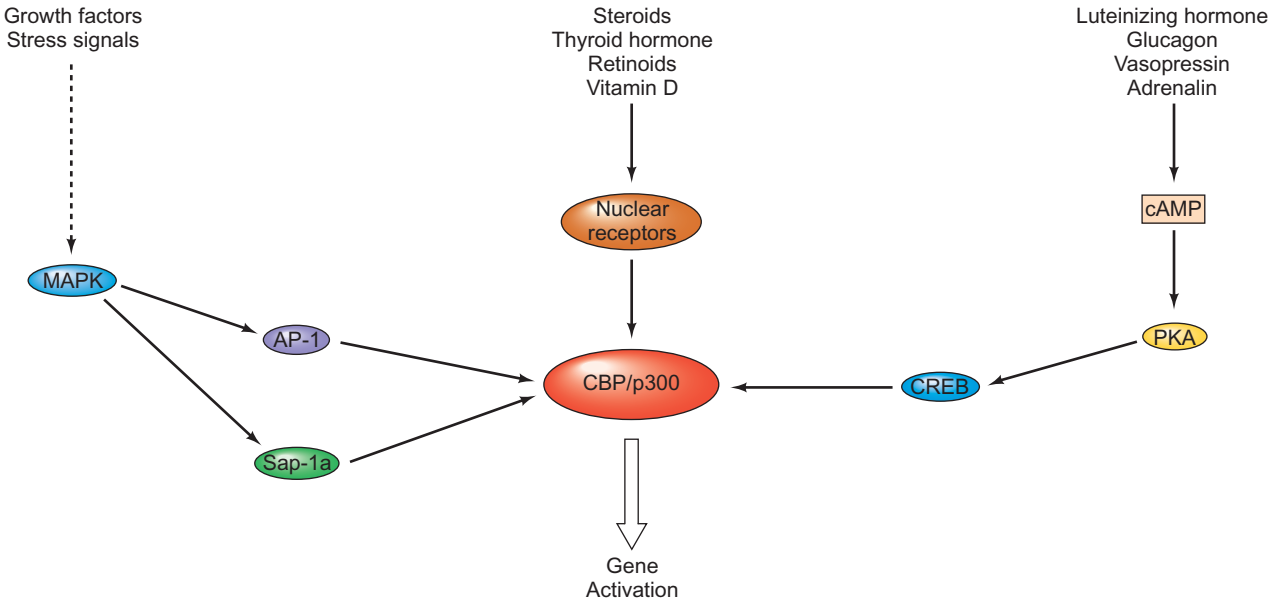


Figure 12.36 Multiple roles of CBP/p300. Three signal transduction pathways that use CBP/p300 to mediate transcription activation are shown converging on CBP/p300 (red) at center. The arrows between pathway members simply indicate position within the pathway (e.g., MAPK acts on AP-1), without indicating the nature of the action (e.g., phosphorylation). This scheme has also been simplified by omitting branches in the pathways. For example MAPK and PKA also phosphorylate nuclear receptors, although the importance of this phosphorylation is unclear. (*Source:* Adapted from Jankneht, R. and T. Hunter, Transcription: A growing coactivator network. *Nature* 383:23, 1996. Copyright © 1996.)

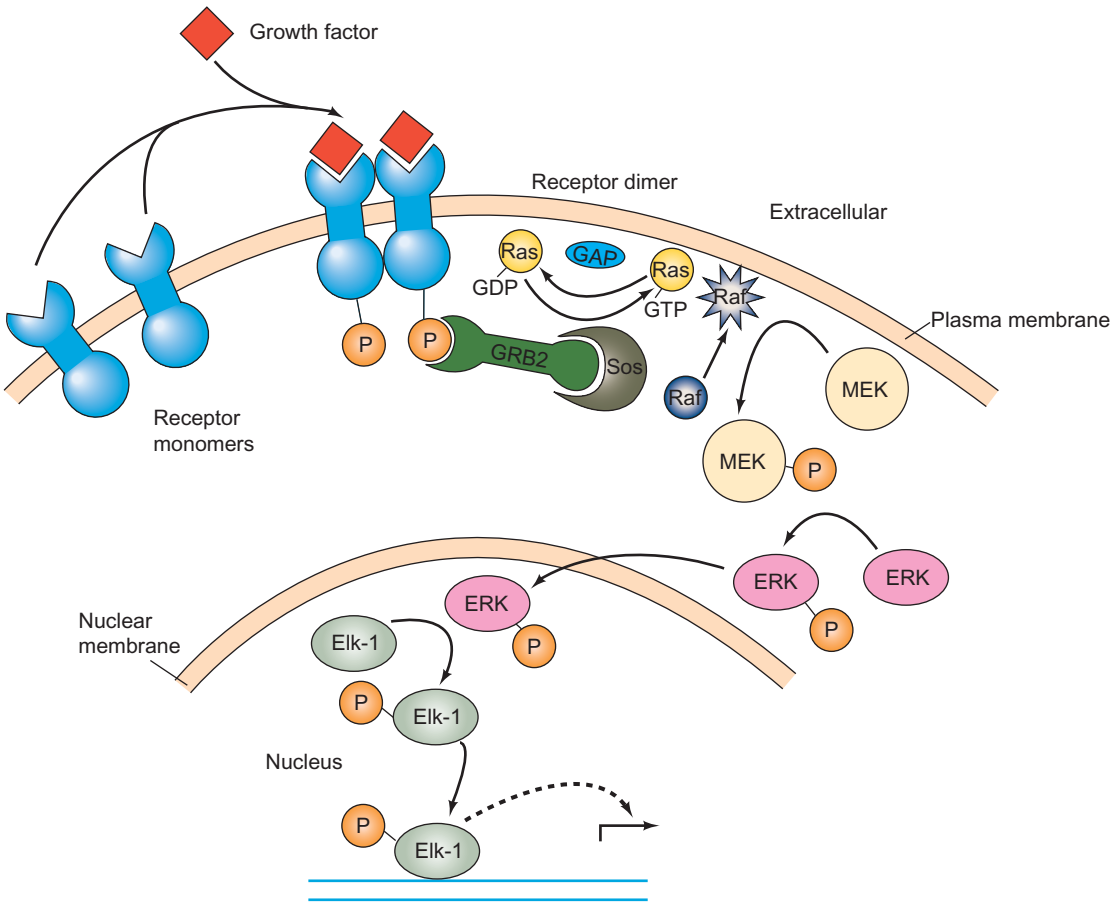


Figure 12.37 Signal transduction pathway involving Ras and Raf. Signal transduction begins (top) when a growth factor or other extracellular signaling molecule (red) binds to its receptor (blue). In this case, the receptor dimerizes on binding its ligand. The intracellular protein tyrosine kinase domain of each receptor monomer then phosphorylates its partner. The new phosphotyrosines can then be recognized by an adapter molecule called GRB2 (dark green), which in turn binds to the Ras exchanger Sos. Sos (gray) is activated to replace GDP on Ras with GTP, thus activating Ras (yellow). Ras delivers Raf (purple) to the cell membrane, where Raf becomes activated. The protein serine/threonine kinase domain of Raf is activated at the membrane, so it phosphorylates MAPK/ERK kinase (MEK, pale yellow), which phosphorylates extracellular-signal-regulated kinase (ERK, pink), which enters the nucleus and phosphorylates the transcription factor Elk-1 (light green). This activates Elk-1, which stimulates transcription of certain genes. The end result is more rapid cell division.

and the nuclear receptor pathway. The first two rely heavily on protein phosphorylation cascades to activate members of the pathway and ultimately to activate transcription. Let us explore the Ras–Raf pathway in more detail and see how aberrant members of the pathway can lead a cell to lose control over its growth and become a cancer cell.

Figure 12.37 presents a Ras–Raf pathway with mammalian names for the proteins. The same pathway operates in other organisms (famously in *Drosophila*) where the proteins have different names. The pathway begins when an extracellular agent, such as a growth factor, interacts with a receptor in the cell membrane. The agent (**epidermal growth factor [EGF],** for example) binds to the extracellular domain of its receptor. This binding stimulates two adjacent receptors to come together to form a dimer, causing the intracellular domains, which have protein tyrosine kinase activity, to phosphorylate each other. Notice how the transmembrane receptor has transduced the signal across the cell membrane into the cell (Latin, *transducere,* meaning "to lead across"). Once the intracellular domains of the receptors are phosphorylated, the new phosphotyrosines attract adapter proteins such as **GRB2** (pronounced "grab two") that have specialized phosphotyrosine binding sites called **SH2 domains.** These are named for similar sites on an oncoprotein called $pp60^{src}$, which can transform cells from normal to tumor-like behavior; SH stands for "Src homology." GRB2 has another domain called **SH3** (also found in $pp60^{src}$) that attracts proteins with a particular kind of hydrophobic α-helix, such as **Sos.** Sos is a **Ras exchanger** that can replace GDP on the protein **Ras** with GTP, thereby activating the Ras protein. Ras contains an endogenous GTPase activity that can hydrolyze the GTP to GDP, inactivating the Ras protein. This GTPase activity is very weak by itself, but it can

be strongly stimulated by another protein called **GTPase activator protein (GAP).** Thus, GAP is an inhibitor of this signal transduction pathway.

Once activated, Ras attracts another protein, **Raf,** to the inner surface of the cell membrane, where Raf is activated. Raf is another protein kinase, but it adds phosphate groups to serines rather than to tyrosines. Its target is another protein serine kinase called **MEK (MAPK/ERK kinase).** In turn, MEK phosphorylates another protein kinase known as **ERK (extracellular-signal-regulated kinase),** activating it. Activated ERK can then phosphorylate a variety of cytoplasmic proteins, and it can also move into the nucleus, where it phosphorylates, and thereby activates, several activators, including **Elk-1.** Activated Elk-1 then stimulates transcription of genes whose products promote cell division.

Thus, one signal transduction pathway that begins with a growth factor interacting with the surface of a cell and ends with enhanced transcription of growth-promoting genes, can be pictured as follows:

Growth factor→receptor→GRB2→Sos→Ras→Raf→MEK→ERK→Elk-1→enhanced transcription→more cell division

It is not surprising that the genes encoding many of the carriers in this pathway are **proto-oncogenes,** whose mutation can lead to runaway cell growth and cancer. If these genes overproduce their products, or make products that are hyperactive, the whole pathway can speed up, leading to abnormally enhanced cell growth and, ultimately, to cancer.

Notice the amplifying power of this pathway. One molecule of EGF can lead to the activation of many molecules of Ras, each of which can activate many molecules of Raf. And, because Raf and the kinases that follow it in the pathway are all enzymes, each can activate many molecules of the next member of the pathway. By the end, one molecule of EGF can yield a great number of activated transcription factors, leading to a burst of new transcription. We should also note that this is only one pathway leading through Ras. In reality, the pathway branches at several points, rather like a web. This kind of interaction between members of different signal transduction pathways is called **cross talk.**

SUMMARY Signal transduction pathways usually begin with a signaling molecule that interacts with a receptor on the cell surface, which sends the signal into the cell, and frequently leads to altered gene expression. Many signal transduction pathways, including the Ras–Raf pathway, rely on protein phosphorylation to pass the signal from one protein to another. This amplifies the signal at each step.

SUMMARY

Eukaryotic activators are composed of at least two domains: a DNA-binding domain and a transcription-activating domain. DNA-binding domains include motifs such as a zinc module, homeodomain, bZIP, or bHLH motif. Transcription-activating domains can be acidic, glutamine-rich, or proline-rich.

Zinc fingers are composed of an antiparallel β-sheet, followed by an α-helix. The β-sheet contains two cysteines, and the α-helix two histidines, that are coordinated to a zinc ion. This coordination of amino acids to the metal helps form the finger-shaped structure. The specific recognition between the finger and its DNA target occurs in the major groove.

The DNA-binding motif of the GAL4 protein contains six cysteines that coordinate two zinc ions in a bimetal thiolate cluster. This DNA-binding motif contains a short α-helix that protrudes into the DNA major groove and makes specific interactions there. The GAL4 monomer also contains an α-helical dimerization motif that forms a parallel coiled coil with the α-helix on the other GAL4 monomer.

Type I nuclear receptors reside in the cytoplasm, bound to another protein. When they bind their hormone ligands, these receptors release their cytoplasmic partners, move to the nucleus, bind to enhancers, and thereby act as activators. The glucocorticoid receptor is representative of this group. It has a DNA-binding domain containing two zinc modules. One module contains most of the DNA-binding residues (in a recognition α-helix), and the other module provides the surface for protein–protein interaction to form a dimer. These zinc modules use four cysteine residues to complex the zinc ion, instead of two cysteines and two histidines as seen in classical zinc fingers.

The homeodomains in eukaryotic activators contain a DNA-binding motif that functions in much the same way as prokaryotic helix-turn-helix motifs, where a recognition helix fits into the DNA major groove and contacts specific residues there.

The bZIP proteins dimerize through a leucine zipper, which puts the adjacent basic regions of each monomer in position to embrace the DNA target site like a pair of tongs. Similarly, the bHLH proteins dimerize through a helix-loop-helix motif, which allows the basic parts of each long helix to grasp the DNA target site, much as the bZIP proteins do. The bHLH and bHLH-ZIP domains bind to DNA in the same way, but the latter have extra dimerization potential due to their leucine zippers.

The DNA-binding and transcription-activation domains of activator proteins are independent modules. Hybrid proteins with the DNA-binding domain of one protein and the transcription-activation domain of another still function as activators.

Activators function by contacting general transcription factors and stimulating the assembly of preinitiation complexes at promoters. For class II promoters, this assembly may occur by stepwise buildup of the general transcription factors and RNA polymerase II, as observed in vitro, or it may occur by recruitment of a large holoenzyme that includes RNA polymerase and most of the general transcription factors. Additional factors (perhaps just TBP or TFIID) may be recruited independently of the holoenzyme.

Dimerization is a great advantage to an activator because it increases the affinity between the activator and its DNA target. Some activators form homodimers, but others function better as heterodimers.

The essence of enhancer function—protein–protein interaction between activators bound to the enhancers, and general transcription factors and RNA polymerase bound to the promoter—seems in many cases to be mediated by looping out the DNA in between. At least in theory, this can also account for the effects of multiple enhancers on gene transcription. DNA looping could bring the activators bound to each enhancer close to the promoter where they could stimulate transcription, perhaps in a cooperative way.

Transcription appears to be concentrated in transcription factories within the nucleus, where an average of about 14 polymerases II and III are active. The existence of transcription factories implies the existence of DNA loops between genes being transcribed in the same factory.

Complex enhancers enable a gene to respond differently to different combinations of activators. This arrangement gives cells exquisitely fine control over their genes in different tissues, or at different times in a developing organism.

The architectural transcription factor LEF-1 binds to the minor groove of its DNA target through its HMG domain and induces strong bending in the DNA. LEF-1 does not enhance transcription by itself, but the bending it induces probably helps other activators bind and interact with still other activators and the general transcription factors to stimulate transcription.

An enhanceosome is a nucleoprotein complex containing a collection of activators bound to an enhancer so as to stimulate transcription. The archetypical enhanceosome involves the IFN-β enhancer. Its structure involves eight polypeptides bound cooperatively to an essentially straight 55-bp stretch of DNA. HMGA1a is essential for this cooperative binding, but it is not part of the final enhanceosome.

Insulators are DNA elements that can shield genes from activation by enhancers (enhancer-blocking activity) or repression by silencers (barrier activity). Some insulators have both enhancer blocking and barrier activities, but some have only one or the other. Insulators may do their job by working in pairs that bind proteins that can interact to form DNA loops. These loops would isolate enhancers and silencers so they can no longer stimulate or repress promoters. In this way, insulators may establish boundaries between DNA regions in a chromosome. Two or more insulators between an enhancer and a promoter cancel each other's effect, perhaps by binding proteins that interact with each other and thereby prevent the DNA looping that would isolate the enhancer from the promoter. Alternatively, the interaction between adjacent insulator-binding proteins could prevent the association of the insulators with insulator bodies, and this could block insulator activity.

Several different activators, including CREB, the nuclear receptors, and AP-1, do not activate transcription by contacting the basal transcription apparatus directly. Instead, upon being phosphorylated, they contact a coactivator called CBP (or its homolog p300), which in turn contacts the basal transcription apparatus and recruits it to promoters. CBP/p300 bound to nuclear receptor-response elements can also recruit CARM1, which methylates an arginine on CBP/p300 required to interact with CREB. This prevents activation of CREB-responsive genes.

Some activators and coactivators are controlled by ubiquitin-mediated destruction. The proteins are ubiquitylated, which marks them for destruction by the 26S proteasome. Ubiquitylation (especially monoubiquitylation) of some activators can have an activating effect, but polyubiquitylation marks these same proteins for destruction. Proteins from the 19S regulatory particle of the proteasome can stimulate transcription, perhaps by remodeling and thereby activating transcription factors.

Some activators can be sumoylated (coupled to a small, ubiquitin-like protein called SUMO), which causes them to be sequestered in nuclear bodies, where they cannot carry out their transcription activation function. Nonhistone activators and repressors can be acetylated by HATs, and this acetylation can have either positive or negative effects.

Signal transduction pathways usually begin with a signaling molecule that interacts with a receptor on the cell surface, which sends the signal into the cell, and frequently leads to altered gene expression. Many signal transduction pathways, including the Ras–Raf pathway, rely on protein phosphorylation to pass the signal from one protein to another. This enzymatic action amplifies the signal at each step. However, ubiquitylation and sumoylation of activators and other signal transduction pathway members can also play major roles in these pathways.

REVIEW QUESTIONS

1. List three different classes of DNA-binding domains found in eukaryotic transcription factors.
2. List three different classes of transcription-activation domains in eukaryotic transcription factors.

3. Draw a diagram of a zinc finger. Point out the DNA-binding motif of the finger.
4. List one important similarity and three differences between a typical prokaryotic helix-turn-helix domain and the Zif268 zinc finger domain.
5. Draw a diagram of the dimer composed of two molecules of the N-terminal 65 amino acids of the GAL4 protein, interacting with DNA. Your diagram should show clearly the dimerization domains and the motifs in the two DNA-binding domains interacting with their DNA-binding sites. What metal ions and coordinating amino acids, and how many of each, are present in each DNA-binding domain?
6. In general terms, what is the function of a nuclear receptor?
7. Explain the difference between type I and II nuclear receptors and give an example of each.
8. What metal ions and coordinating amino acids, and how many of each, are present in each DNA-binding domain of a nuclear receptor? What part of the DNA-binding domain contacts the DNA bases?
9. What is the nature of the homeodomain? What other DNA-binding domain does it most resemble?
10. Draw a diagram of a leucine zipper seen from the end. How does this diagram illustrate the relationship between the structure and function of the leucine zipper?
11. Draw a diagram of a bZIP protein interacting with its DNA-binding site.
12. Describe and show the results of an experiment that illustrates the independence of the DNA-binding and transcription-activating domains of an activator.
13. Present two models of recruitment of the class II preinitiation complex, one involving a holoenzyme, the other not.
14. Describe and give the results of an experiment that shows that an acidic transcription-activating domain binds to TFIID.
15. Present evidence that favors the holoenzyme recruitment model.
16. Present two lines of evidence that argue against the holoenzyme recruitment model.
17. Why is a protein dimer (or tetramer) so much more effective than a monomer in DNA binding? Why is it important for a transcription activator to have a high affinity for specific sequences in DNA?
18. Present three models to explain how an enhancer can act on a promoter hundreds of base pairs away.
19. Describe and give the results of an experiment that shows the effect of isolating an enhancer on a separate circle of DNA intertwined with another circle of DNA that contains the promoter. Which model(s) of enhancer activity does this experiment favor? Why?
20. Describe how you would perform a hypothetical 3C experiment. Describe the results you would get, and give an interpretation.
21. What advantage do complex enhancers confer on a gene?
22. Describe how you would identify transcription factories in a cell nucleus. Why are both in vitro and in vivo transcription essential parts of the procedure? Why does the existence of transcription factories imply that chromatin loops occur in the nucleus?
23. LEF-1 is an activator of the human T-cell receptor α-chain, yet LEF-1 by itself does not activate this gene. How does LEF-1 act? Describe and show the results of an experiment that supports your answer.
24. Does LEF-1 bind in the major or minor groove of its DNA target? Present evidence to support your answer.
25. What do insulators do?
26. Diagram a model to explain the following results: (a) Having one insulator between an enhancer and a promoter partially blocks enhancer activity. (b) Having two insulators between an enhancer and a promoter does not block enhancer activity. (c) Having one insulator on either side of an enhancer strongly blocks enhancer activity.
27. What is the effect of three copies of an insulator between an enhancer and a promoter? How do you explain this phenomenon?
28. Present evidence for the hypothesis that an insulator blocks enhancement by interacting with nearby enhancers and promoters. What are the difficulties in generalizing this hypothesis to all insulators?
29. Describe and give the results of an experiment that shows the effects of Mediator.
30. Draw diagrams to illustrate the action of CBP as a coactivator of (a) phosphorylated CREB; (b) a nuclear receptor.
31. How do signal transduction pathways amplify their signals? Present an example.
32. Present a hypothesis to explain the negative effect of ubiquitin on transcription.
33. Present a hypothesis to explain the positive effect of proteasome proteins on transcription.

ANALYTICAL QUESTIONS

1. Design an experiment to show that TFIID binds directly to an acidic activating domain. Show sample positive results.
2. You are studying the human *BLU* gene, which is under the control of three enhancers. You suspect that the proteins that bind to these enhancers interact with each other to form an enhanceosome that is required for activation. What spacing among these enhancers is optimal for such interaction? What changes in this spacing could you introduce to test your hypothesis? What results would you expect?
3. Consider Figure 12.22a. What primers would you use in a 3C experiment to show association between the *ICR* insulator and each of the *Igf2* promoters *P1, P2,* and *P3,* on the maternal chromosome.
4. You are going to create a human activator (eA1) that controls a set of genes responsible for academic success. You aim to create an activator that includes the components deemed essential through the study of other activators.

What is the composition of your activator? In the process you also create two additional distinct activators (eA2, eA3). What experiments would you run to determine which activator works best? Suppose you wanted the activator to work in women students but not in men. How might you arrange that? What kind of activator would you have to design to make that work?

SUGGESTED READINGS

General References and Reviews

Bell, A.C. and G. Felsenfeld. 1999. Stopped at the border: boundaries and insulators. *Current Opinion in Genetics and Development* 9:191–98.

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

Carey, M. 1994. Simplifying the complex. *Nature* 368:402–3.

Conaway, R.C., C.S. Brower, and J.W. Conaway. 2002. Emerging roles of ubiquitin in transcription regulation. *Science* 296:1254–58.

Freiman, R.N. and R. Tjian. 2003. Regulating the regulators: Lysine modifications make their mark. *Cell* 112:11–17.

Goodrich, J.A., G. Cutler, and R. Tjian. 1996. Contacts in context: Promoter specificity and macromolecular interactions in transcription. *Cell* 84:825–30.

Hampsey, M. and D. Reinberg. 1999. RNA polymerase II as a control panel for multiple coactivator complexes. *Current Opinion in Genetics and Development* 9:132–39.

Janknecht, R. and T. Hunter. 1996. Transcription. A growing coactivator network. *Nature* 383:22–23.

Montminy, M. 1997. Something new to hang your hat on. *Nature* 387:654–55.

Myer, V.E. and R.A. Young. 1998. RNA polymerase II holoenzymes and subcomplexes. *Journal of Biological Chemistry* 273:27757–60.

Nordheim, A. 1994. CREB takes CBP to tango. *Nature* 370:177–78.

Ptashne, M. and A. Gann. 1997. Transcriptional activation by recruitment. *Nature* 386:569–77.

Roush, W. 1996. "Smart" genes use many cues to set cell fate. *Science* 272:652–53.

Sauer, F. and R. Tjian. 1997. Mechanisms of transcription activation: Differences and similarities between yeast, *Drosophila,* and man. *Current Opinion in Genetics and Development* 7:176–81.

Wallace, J.A. and G. Felsenfeld. 2007. We gather together: insulators and genome organization. *Current Opinion in Genetics and Development* 17:400–407.

Werner, M.H. and S.K. Burley. 1997. Architectural transcription factors: Proteins that remodel DNA. *Cell* 88:733–36.

West, A.G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. *Genes and Development* 16:271–88.

Research Articles

Barberis, A., J. Pearlberg, N. Simkovich, S. Farrell, P. Reinagle, C. Bamdad, G. Sigal, and M. Ptashne. 1995. Contact with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* 81:359–68.

Borggrefe, T., R. Davis, A. Bareket-Samish, and R.D. Kornberg. 2001. Quantitation of the RNA polymerase II transcription machinery in yeast. *Journal of Biological Chemistry* 276:47150–53.

Brent, R. and M. Ptashne. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 43:729–36.

Cai, H.N. and P. Shen 2001. Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* 291:493–95.

Dunaway, M. and P. Dröge. 1989. Transactivation of the *Xenopus* rRNA gene promoter by its enhancer. *Nature* 341:657–59.

Ellenberger, T.E., C.J. Brandl, K. Struhl, and S.C. Harrison. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein–DNA complex. *Cell* 71:1223–37.

Flanagan, P.M., R.J. Kelleher, 3rd, M.H. Sayre, H. Tschochner, and R.D. Kornberg. 1991. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350:436–38.

Geise, K., J. Cox, and R. Grosschedl. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* 69:185–95.

Jackson, D.A., F.J. Iborra, E.M.M. Manders, and P.R. Cook. 1998. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Molecular Biology of the Cell* 9:1523–36.

Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: A framework for understanding homeodomain–DNA interactions. *Cell* 63:579–90.

Koleske, A.J. and R.A. Young. 1994. An RNA polymerase II holoenzyme responsive to activators. *Nature* 368:466–69.

Krebs, J.E. and Dunaway, M. 1998. The scs and scs′ elements impart a *cis* requirement on enhancer–promoter interactions. *Molecular Cell* 1:301–08.

Lee, M.S. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635–37.

Leuther, K.K., J.M. Salmeron, and S.A. Johnston. 1993. Genetic evidence that an activation domain of GAL4 does not require acidity and may form a β sheet. *Cell* 72:575–85.

Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352:497–505.

Ma, P.C.M., M.A. Rould, H. Weintraub, and C.O. Pabo. 1994. Crystal structure of MyoD bHLH domain–DNA complex: Perspectives on DNA recognition and implications for transcription activation. *Cell* 77:451–59.

Marmorstein, R., M. Carey, M. Ptashne, and S.C. Harrison. 1992. DNA recognition by GAL4: Structure of a protein–DNA complex. *Nature* 356:408–14.

Muravyova, E., A. Golovnin, E. Gracheva, A. Parshikov, T. Belenkaya, V. Pirrotta, and P. Georgiev. 2001. Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* 291:495–98.

O'Shea, E.K., J.D. Klemm, P.S. Kim, and T. Alber. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254:539–44.

Panne, D., T. Maniatis, and S.C. Harrison. 2007. An atomic model of the interferon-β enhanceosome. *Cell* 129:1111–23.

Pavletich, N.P. and C.O. Pabo. 1991. Zinc finger–DNA recognition: Crystal structure of a Zif 268–DNA complex at 2.1 Å. *Science* 252:809–17.

Romano, L.A. and G.A. Wray. 2003. Conservation of *Endo 16* expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. *Development* 130:4187–99.

Ryu, L., L. Zhou, A.G. Ladurner, and R. Tjian. 1999. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* 397:446–50.

Stringer, K.F., C. J. Ingles, and J. Greenblatt. 1990. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* 345:783–86.

Xu, W., H. Chen, K. Du, H. Asahara, M. Tini, B.M. Emerson, M. Montminy, and R.M Evans. 2001. A transcriptional switch mediated by cofactor methylation. *Science* 294:2507–11.

Yoon, Y.S., S. Jeong, Q. Rong, K.-Y. Park, J.H. Chung, and K. Pfeifer. 2007. Analysis of the *H19ICR* insulator. *Molecular and Cellular Biology* 27:3499–3510.

Yuh, C.-H., B. Hamid, and E.H. Davidson. 1998. Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* 279:1896–1902.

Zhu, J. and M. Levine. 1999. A novel *cis*-regulatory element, the PTS, mediates an antiinsulator activity in the *Drosophila* embryo. *Cell* 99:567–75.