ACTIVATING TRANSCRIPTION

5.22.1 Introduction

Key Concepts

- Eukaryotic gene expression is usually controlled at the level of initiation of transcription.

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure

↓

Initiation of transcription

↓

Processing the transcript

↓

Transport to cytoplasm

↓

Translation of mRNA

As we see in**Figure 22.1**, gene expression in eukaryotes is largely controlled at the initiation of transcription. For most genes, this is the major control point in their expression. It involves changes in the structure of chromatin at the promoter (see *Molecular Biology 5.23.11 Promoter activation involves an ordered series of events*), accompanied by the binding of the basal transcription apparatus (including RNA polymerase II) to the promoter. (Regulation at subsequent stages of transcription is rare in eukaryotic cells. Premature termination occurs at some genes, and is counteracted by a kinase, P-TEFb, but otherwise anti-termination does not seem to be employed.)

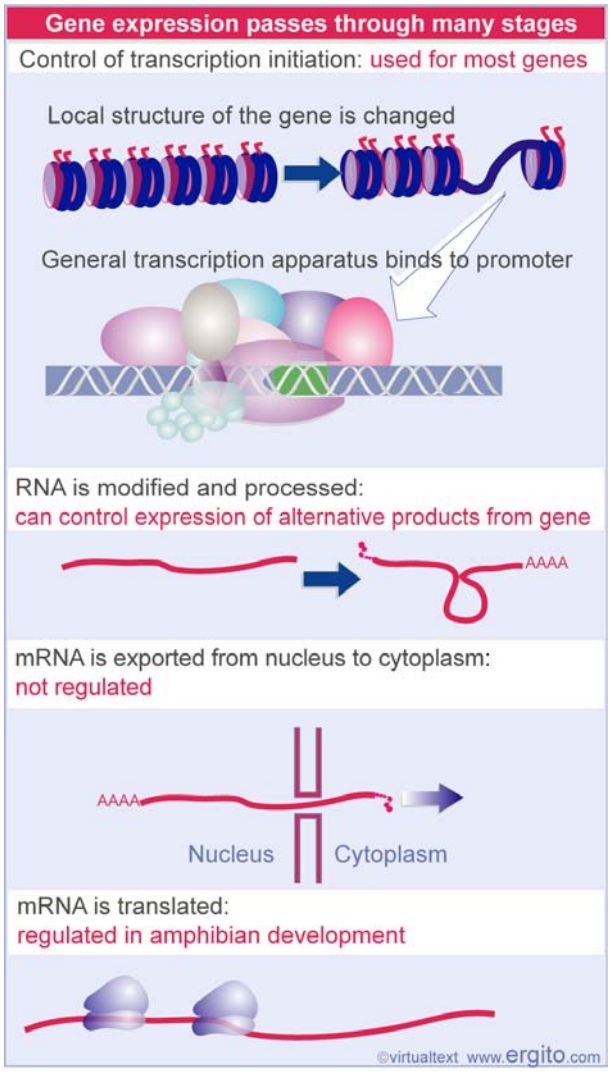

Figure 22.1 Gene expression is controlled principally at the initiation of transcription, and it is rare for the subsequent stages of gene expression to be used to determine whether a gene is expressed, although control of processing may be used to determine which form of a gene is represented in mRNA.

The primary transcript is modified by capping at the 5′ end, and usually also by polyadenylation at the 3′ end. Introns must be excised from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing; some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see *Molecular Biology 5.24.12 Alternative splicing involves differential use of splice junctions*).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it occurs in some embryonic situations. This can involve localization of the

mRNA to specific sites where it is expressed and/or the blocking of initiation of translation by specific protein factors (see *Molecular Biology 6.31.7 How are mRNAs and proteins transported and localized?*).

Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in *Molecular Biology 6.31 Gradients, cascades, and signaling pathways* in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: how does the transcription factor identify its group of target genes; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

ACTIVATING TRANSCRIPTION

5.22.2 There are several types of transcription factors

Key Terms

A **basal factor** is a transcription factor required by RNA polymerase II to form the initiation complex at all promoters. Factors are identified as $TF_{II}X$, where X is a number.

An **activator** is a protein that stimulates the expression of a gene, typically by acting at a promoter to stimulate RNA polymerase. In eukaryotes, the sequence to which it binds in the promoter is called a response element.

A **response element** is a sequence in a eukaryotic promoter or enhancer that is recognized by a specific transcription factor.

Coactivators are factors required for transcription that do not bind DNA but are required for (DNA-binding) activators to interact with the basal transcription factors.

Key Concepts

- The basal apparatus determines the startpoint for transcription.
- Activators determine the frequency of transcription.
- Activators work by making protein-protein contacts with the basal factors.
- Activators may work via coactivators.
- Some components of the transcriptional apparatus work by changing chromatin structure.

Initiation of transcription involves many protein-protein interactions among transcription factors bound at the promoter or at an enhancer as well as with RNA polymerase. We can divide the factors required for transcription into several classes (for review see 1710; 1712). **Figure 22.2** summarizes their properties:

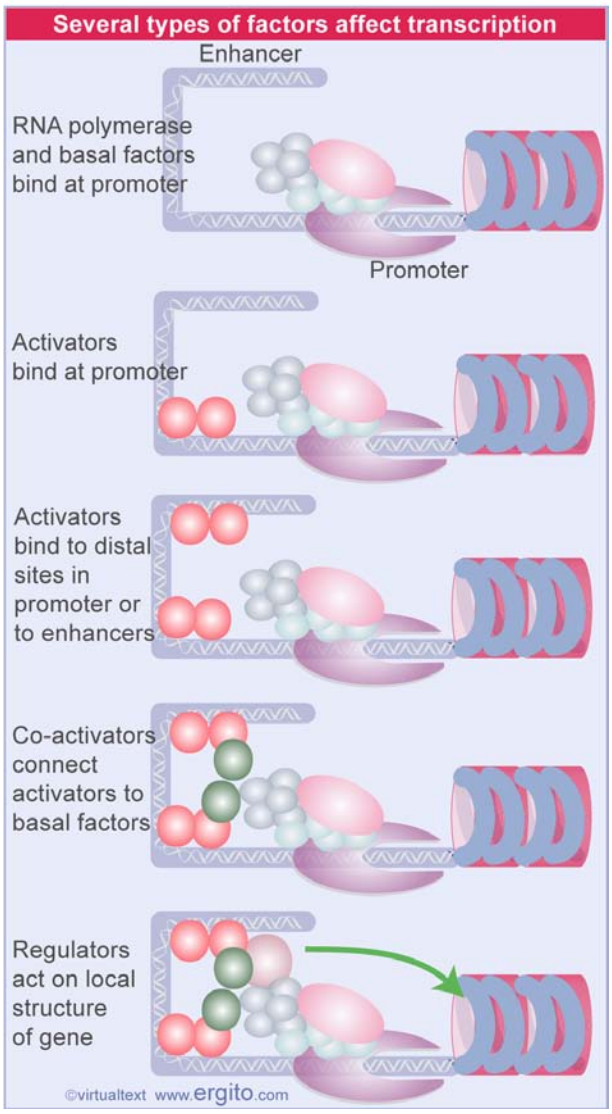

Figure 22.2 Factors involved in gene expression include RNA polymerase and the basal apparatus, activators that bind directly to DNA at the promoter or at enhancers, co-activators that bind to both activators and the basal apparatus, and regulators that act on chromatin structure.

- **Basal factors**, together with RNA polymerase, bind at the startpoint and TATA box (see *Molecular Biology 5.21.10 The basal apparatus assembles at the promoter*).

- **Activators** are transcription factors that recognize specific short consensus elements. They bind to sites in the promoter or in enhancers (see *Molecular Biology 5.21.13 Short sequence elements bind activators*). They act by increasing the efficiency with which the basal apparatus binds to the promoter. They therefore increase the frequency of transcription, and are required for a promoter to function at an adequate level. Some activators act constitutively

(they are ubiquitous), but others have a regulatory role, and are synthesized or activated at specific times or in specific tissues. These factors are therefore responsible for the control of transcription patterns in time and space. The sequences that they bind are called **response elements**.

- Another group of factors necessary for efficient transcription do not themselves bind DNA. **Coactivators** provide a connection between activators and the basal apparatus (see *Molecular Biology 5.22.5 Activators interact with the basal apparatus*). They work by protein-protein interactions, forming bridges between activators and the basal transcription apparatus.

- Some regulators act to make changes in chromatin (see *Molecular Biology 5.23.7 Acetylases are associated with activators*).

The diversity of elements from which a functional promoter may be constructed, and the variations in their locations relative to the startpoint, argues that the activators have an ability to interact with one another by protein-protein interactions in multiple ways. There appear to be no constraints on the potential relationships between the elements. The modular nature of the promoter is illustrated by experiments in which equivalent regions of different promoters have been exchanged. Hybrid promoters, for example, between the thymidine kinase and β-globin genes, work well. This suggests that the main purpose of the elements is to bring the activators they bind into the vicinity of the initiation complex, where protein-protein interactions determine the efficiency of the initiation reaction.

The organization of RNA polymerase II promoters contrasts with that of bacterial promoters, where all the transcription factors must interact directly with RNA polymerase. In the eukaryotic system, only the basal factors interact directly with the enzyme. Activators may interact with the basal factors, or may interact with coactivators that in turn interact with the basal factors. The construction of the apparatus through layers of interactions explains the flexibility with which elements may be arranged, and the distance over which they can be dispersed.

Last updated on 10-1-2001

Reviews

1710. Lemon, B. and Tjian, R. (2000). *Orchestrated response: a symphony of transcription factors for gene control.* Genes Dev. 14, 2551-2569.

1712. Lee, T. I. and Young, R. A. (2000). *Transcription of eukaryotic protein-coding genes.* Annu. Rev. Genet. 34, 77-137.

ACTIVATING TRANSCRIPTION

5.22.3 Independent domains bind DNA and activate transcription

Key Concepts

- DNA-binding activity and transcription-activation are carried by independent domains of an activator.
- The role of the DNA-binding domain is to bring the transcription-activation domain into the vicinity of the promoter.

Activators and other regulatory proteins require two types of ability:

- They recognize specific target sequences located in enhancers, promoters, or other regulatory elements that affect a particular target gene.
- Having bound to DNA, an activator exercises its function by binding to other components of the transcription apparatus.

Can we characterize domains in the activator that are responsible for these activities? Often an activator has separate domains that bind DNA and activate transcription. Each domain behaves as a separate module that functions independently when it is linked to a domain of the other type. The geometry of the overall transcription complex must allow the activating domain to contact the basal apparatus irrespective of the exact location and orientation of the DNA-binding domain.

Upstream promoter elements may be an appreciable distance from the startpoint, and in many cases may be oriented in either direction. Enhancers may be even farther away and always show orientation independence. This organization has implications for both the DNA and proteins. The DNA may be looped or condensed in some way to allow the formation of the transcription complex. And the domains of the activator may be connected in a flexible way, as illustrated diagrammatically in **Figure 22.3**. The main point here is that the DNA-binding and activating domains are independent, and connected in a way that allows the activating domain to interact with the basal apparatus irrespective of the orientation and exact location of the DNA-binding domain.

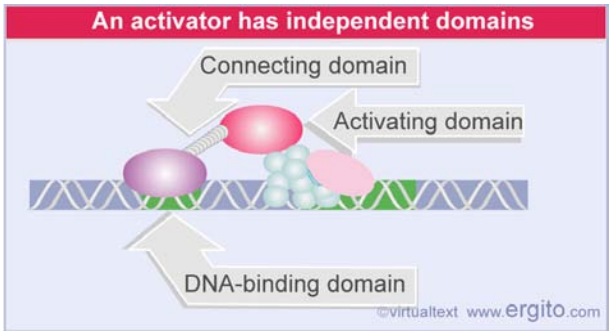

Figure 22.3 DNA-binding and activating functions in a transcription factor may comprise independent domains of the protein.

Binding to DNA is necessary for activating transcription. But does activation depend on the *particular* DNA-binding domain?

Figure 22.4 illustrates an experiment to answer this question. The activator GAL4 has a DNA-binding domain that recognizes a *UAS*, and an activating domain that stimulates initiation at the target promoter. The bacterial repressor LexA has an N-terminal DNA-binding domain that recognizes a specific operator. When LexA binds to this operator, it represses the adjacent promoter. In a "swap" experiment, the DNA-binding domain of LexA can be substituted for the DNA-binding domain of GAL4. The hybrid gene can then be introduced into yeast together with a target gene that contains either the *UAS* or a LexA operator.

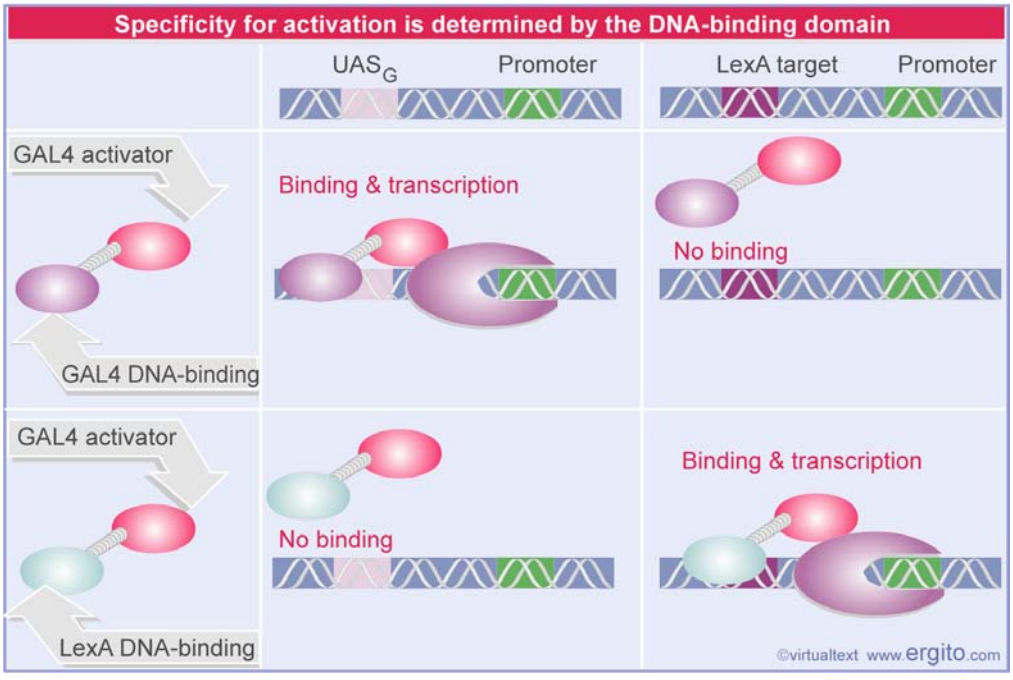

Figure 22.4 The ability of GAL4 to activate transcription is independent of its specificity for binding DNA. When the GAL4 DNA-binding domain is replaced by the LexA DNA-binding domain, the hybrid protein can activate transcription when a LexA operator is placed near a promoter.

An authentic GAL4 protein can activate a target gene only if it has a *UAS*. The LexA repressor by itself of course lacks the ability to activate either sort of target. The LexA-GAL4 hybrid can no longer activate a gene with a *UAS*, but it can now activate a gene that has a LexA operator!

This result fits the modular view of transcription activators. The DNA-binding domain serves to bring the protein into the right location. Precisely how or where it is bound to DNA is irrelevant, but, once it is there, the transcription-activating domain can play its role. According to this view, it does not matter whether the transcription-activating domain is brought to the vicinity of the promoter by recognition of a *UAS* via the DNA-binding domain of GAL4 or by recognition of a LexA operator via the LexA specificity module. The ability of the two types of module to function in hybrid proteins suggests that each domain of the protein folds independently into an active structure that is not influenced by the rest of the protein (for review see 230; 232).

The idea that activators have independent domains that bind DNA and that activate transcription is reinforced by the ability of the tat protein of HIV to stimulate initiation without binding DNA at all. The tat protein binds to a region of secondary structure in the RNA product; the part of the RNA required for tat action is called the *tar* sequence. A model for the role of the tat-*tar* interaction in stimulating transcription is shown in **Figure 22.5**.

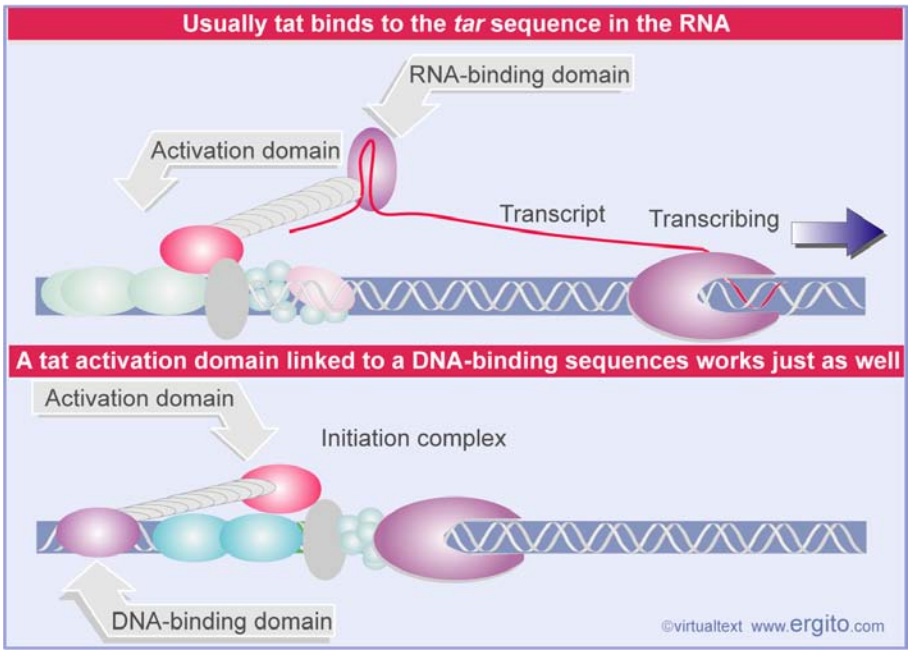

Figure 22.5 The activating domain of the tat protein of HIV can stimulate transcription if it is tethered in the vicinity by binding to the RNA product of a previous round of transcription. Activation is independent of the means of tethering, as shown by the substitution of a DNA-binding domain for the RNA-binding domain.

The *tar* sequence is located just downstream of the startpoint, so that when tat binds to *tar*, it is brought into the vicinity of the initiation complex. This is sufficient to ensure that its activation domain is in close enough proximity to the initiation complex. The activation domain interacts with one or more of the transcription

factors bound at the complex in the same way as an activator. (Of course, the first transcript must be made in the absence of tat in order to provide the binding site.)

An extreme demonstration of the independence of the localizing and activating domains is indicated by some constructs in which tat was engineered so that the activating domain was connected to a DNA-binding domain instead of to the usual *tar*-binding sequence. When an appropriate target site was placed into the promoter, the tat activating-domain could activate transcription. This suggests that we should think of the DNA-binding (or in this case the RNA-binding) domain as providing a "tethering" function, *whose main purpose is to ensure that the activating domain is in the vicinity of the initiation complex.*

The notion of tethering is a more specific example of the general idea that initiation requires a high concentration of transcription factors in the vicinity of the promoter. This may be achieved when activators bind to enhancers in the general vicinity, when activators bind to upstream promoter components, or in an extreme case by tethering to the RNA product. The common requirement of all these situations is flexibility in the exact three dimensional arrangement of DNA and proteins. The principle of independent domains is common in transcriptional activators.

We might view the function of the DNA-binding domain as *bringing the activating domain into the vicinity of the startpoint.* This explains why the exact locations of DNA-binding sites can vary within the promoter.

Reviews

230. Guarente, L. (1987). *Regulatory proteins in yeast*. Annu. Rev. Genet. 21, 425-452.

232. Ptashne, M. (1988). *How eukaryotic transcriptional activators work*. Nature 335, 683-689.

ACTIVATING TRANSCRIPTION

5.22.4 The two hybrid assay detects protein-protein interactions

Key Terms

Two hybrid assay detects interaction between two proteins by means of their ability to bring together a DNA-binding domain and a transcription-activating domain. The assay is performed in yeast using a reporter gene that responds to the interaction.

Key Concepts

- The two hybrid assay works by requiring an interaction between two proteins where one has a DNA-binding domain and the other has a transcription-activation domain.

The model of domain independence is the basis for an extremely useful assay for detecting protein interactions. In effect, we replace the connecting domain in **Figure 22.3** with a protein-protein interaction. The principle is illustrated in **Figure 22.6**. We fuse one of the proteins to be tested to a DNA-binding domain. We fuse the other protein to a transcription-activating domain. (This is done by linking the appropriate coding sequences in each case and making synthetic proteins by expressing each hybrid gene.)

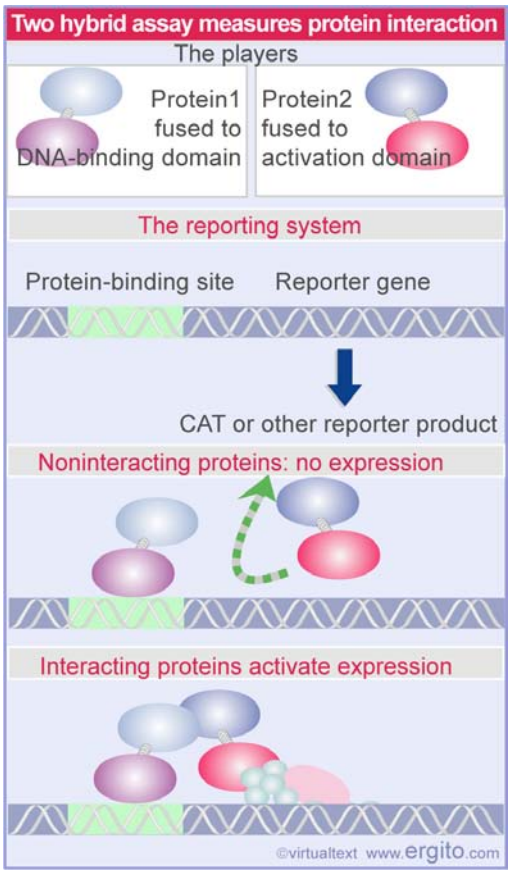

Figure 22.6 The two hybrid technique tests the ability of two proteins to interact by incorporating them into hybrid proteins where one has a DNA-binding domain and the other has a transcription-activating domain.

If the two proteins that are being tested can interact with one another, the two hybrid proteins will interact. This is reflected in the name of the technique: the **two hybrid** assay (952). The protein with the DNA-binding domain binds to a reporter gene that has a simple promoter containing its target site. But it cannot activate the gene by itself. Activation occurs only if the second hybrid binds to the first hybrid to bring the activation domain to the promoter. Any reporter gene can be used where the product is readily assayed, and this technique has given rise to several automated procedures for rapidly testing protein-protein interactions.

The effectiveness of the technique dramatically illustrates the modular nature of proteins. Even when fused to another protein, the DNA-binding domain can bind to DNA and the transcription-activating domain can activate transcription. Correspondingly, the interaction ability of the two proteins being tested is not inhibited by the attachment of the DNA-binding or transcription-activating domains. (Of course, there are some exceptions where these simple rules do not apply and interference between the domains of the hybrid protein prevents the technique from working.)

The power of this assay is that it requires only that the two proteins being tested can

interact with each other. They need not have anything to do with transcription. Because of the independence of the DNA-binding and transcription-activating domains, all we require is that they are brought together. This will happen so long as the two proteins being tested can interact in the environment of the nucleus.

Last updated on 4-30-2000

References

952. Fields, S., and Song, O. (1989). *A novel genetic system to detect protein-protein interactions.* Nature 340, 245-246.

ACTIVATING TRANSCRIPTION

5.22.5 Activators interact with the basal apparatus

Key Terms

Mediator is a large protein complex associated with yeast bacterial RNA polymerase II. It contains factors that are necessary for transcription from many or most promoters.

Key Concepts

- The principle that governs the function of all activators is that a DNA-binding domain determines specificity for the target promoter or enhancer.
- The DNA-binding domain is responsible for localizing a transcription-activating domain in the proximity of the basal apparatus.
- An activator that works directly has a DNA-binding domain and an activating domain.
- An activator that does not have an activating domain may work by binding a coactivator that has an activating domain.
- Several factors in the basal apparatus are targets with which activators or coactivators interact.
- RNA polymerase may be associated with various alternative sets of transcription factors in the form of a holoenzyme complex.

An activator may work directly when it consists of a DNA-binding domain linked to a transcription-activating domain, as illustrated in **Figure 22.3**. In other cases, the activator does not itself have a transcription-activating domain, but binds another protein – a coactivator – that has the transcription-activating domain. **Figure 22.8** shows the action of such an activator. We may regard coactivators as transcription factors whose specificity is conferred by the ability to bind to DNA-binding transcription factors instead of directly to DNA. A particular activator may require a specific coactivator.

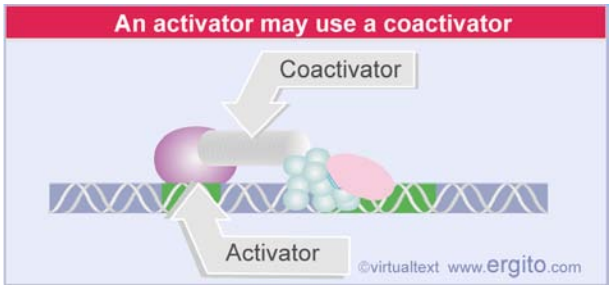

Figure 22.8 An activator may bind a coactivator that contacts the basal apparatus.

But although the protein components are organized differently, the mechanism is the same. An activator that contacts the basal apparatus directly has an activation domain covalently connected to the DNA-binding domain. When an activator works through a coactivator, the connections involve noncovalent binding between protein subunits (compare **Figure 22.3** and **Figure 22.8**). The same interactions are responsible for activation, irrespective of whether the various domains are present in the same protein subunit or divided into multiple protein subunits (645).

A transcription-activating domain works by making protein-protein contacts with general transcription factors that promote assembly of the basal apparatus (for review see 218; 221). Contact with the basal apparatus may be made with any one of several basal factors, typically $TF_{II}D$, $TF_{II}B$, or $TF_{II}A$. All of these factors participate in early stages of assembly of the basal apparatus (see **Figure 21.15**). **Figure 22.9** illustrates the situation when such a contact is made. The major effect of the activators is to influence the assembly of the basal apparatus (646; 659).

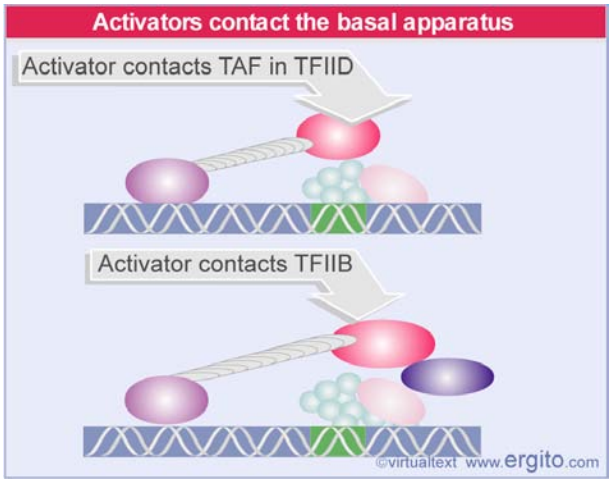

Figure 22.9 Activators may work at different stages of initiation, by contacting the TAFs of TFIID or contacting TFIIB.

$TF_{II}D$ may be the most common target for activators, which may contact any one of several TAFs. In fact, a major role of the TAFs is to provide the connection from the basal apparatus to activators. This explains why TBP alone can support basal-level transcription, but the TAFs of $TF_{II}D$ are required for the higher levels of transcription that are stimulated by activators. Different TAFs in $TF_{II}D$ may provide surfaces that interact with different activators. Some activators interact only with individual TAFs; others interact with multiple TAFs. We assume that the interaction either assists binding of $TF_{II}D$ to the TATA box or assists the binding of other activators around the $TF_{II}D$-TATA box complex. In either case, the interaction stabilizes the basal transcription complex; this speeds the process of initiation, and thereby increases use of the promoter.

The activating domains of the yeast activators GAL4 and GCN4 have multiple negative charges, giving rise to their description as "acidic activators." Another particularly effective activator of this type is carried by the VP16 protein of the Herpes Simplex Virus. (VP16 does not itself have a DNA-binding domain, but interacts with the transcription apparatus via an intermediary protein.) Experiments

to characterize acidic activator functions have often made use of the VP16 activating region linked to a DNA-binding motif (658).

Acidic activators function by enhancing the ability of $TF_{II}B$ to join the basal initiation complex. Experiments *in vitro* show that binding of $TF_{II}B$ to an initiation complex at an adenovirus promoter is stimulated by the presence of GAL4 or VP16 acid activators; and the VP16 activator can bind directly to $TF_{II}B$. Assembly of $TF_{II}B$ into the complex at this promoter is therefore a rate-limiting step that is stimulated by the presence of an acidic activator.

The resilience of an RNA polymerase II promoter to the rearrangement of elements, and its indifference even to the particular elements present, suggests that the events by which it is activated are relatively general in nature. Any activators whose activating region is brought within range of the basal initiation complex may be able to stimulate its formation. Some striking illustrations of such versatility have been accomplished by constructing promoters consisting of new combinations of elements. For example, when a yeast UAS_G element is inserted near the promoter of a higher eukaryotic gene, this gene can be activated by GAL4 in a mammalian cultured cell. Whatever means GAL4 uses to activate the promoter seems therefore to have been conserved between yeast and higher eukaryotes. The GAL4 protein must recognize some feature of the mammalian transcription apparatus that resembles its normal contacts in yeast.

How does an activator stimulate transcription? We can imagine two general types of model:

- The recruitment model argues that its sole effect is to increase the binding of RNA polymerase to the promoter.
- An alternative model is to suppose that it induces some change in the transcriptional complex, for example, in the conformation of the enzyme, which increases its efficiency.

A test of these models in one case in yeast showed that recruitment can account for activation. When the concentration of RNA polymerase was increased sufficiently, the activator failed to produce any increase in transcription, suggesting that its sole effect is to increase the effective concentration of RNA polymerase at the promoter.

Adding up all the components required for efficient transcription – basal factors, RNA polymerase, activators, coactivators – we get a very large apparatus, consisting of >40 proteins. Is it feasible for this apparatus to assemble step by step at the promoter? Some activators, coactivators, and basal factors may assemble stepwise at the promoter, but then may be joined by a very large complex consisting of RNA polymerase preassembled with further activators and coactivators, as illustrated in **Figure 22.7** (for review see 1710).

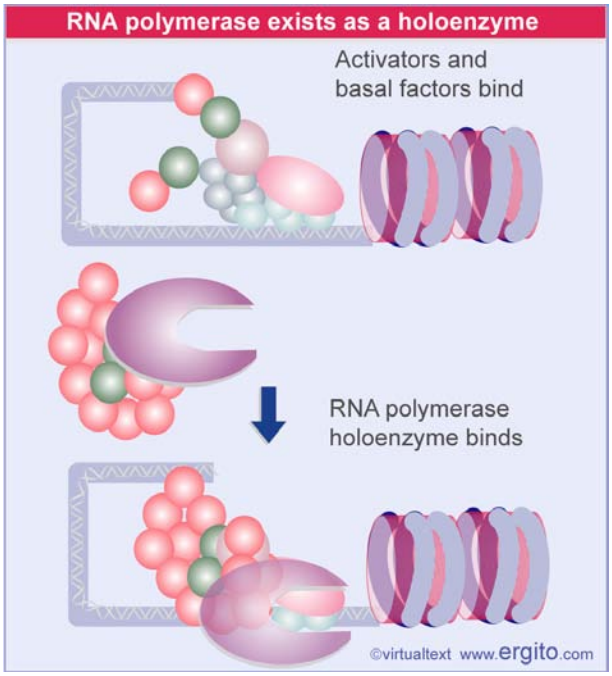

Figure 22.7 RNA polymerase exists as a holoenzyme containing many activators.

Several forms of RNA polymerase have been found in which the enzyme is associated with various transcription factors. The most prominent "holoenzyme complex" in yeast (defined as being capable of initiating transcription without additional components) consists of RNA polymerase associated with a 20-subunit complex called **mediator** (1713; for review see 1711). The mediator includes products of several genes in which mutations block transcription, including some *SRB* loci (so named because many of their genes were originally identified as suppressors of mutations in *R*NA polymerase *B.*) The name was suggested by its ability to mediate the effects of activators. Mediator is necessary for transcription of most yeast genes. Homologous complexes are required for the transcription of most higher eukaryotic genes (2411). Mediator undergoes a conformational change when it interacts with the CTD domain of RNA polymerase (2412). It can transmit either activating or repressing effects from upstream components to the RNA polymerase. It is probably released when a polymerase starts elongation. Some transcription factors influence transcription directly by interacting with RNA polymerase or the basal apparatus, but others work by manipulating structure of chromatin (see *Molecular Biology 5.23.3 Chromatin remodeling is an active process*).

Last updated on 4-22-2002

Reviews

218. Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987). *Regulation of inducible and tissue-specific gene expression.* Science 236, 1237-1245.

221. Mitchell, P., and Tjian, R. (1989). *Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins*. Science 245, 371-378.

1710. Lemon, B. and Tjian, R. (2000). *Orchestrated response: a symphony of transcription factors for gene control.* Genes Dev. 14, 2551-2569.

1711. Myers, L. C. and Kornberg, R. D. (2000). *Mediator of transcriptional regulation.* Annu. Rev. Biochem. 69, 729-749.

References

645. Pugh, B. F. and Tjian, R. (1990). *Mechanism of transcriptional activation by Sp1: evidence for coactivators*. Cell 61, 1187-1197.

646. Dynlacht, B. D., Hoey, T., and Tjian, R. (1991). *Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation*. Cell 66, 563-576.

658. Ma, J. and Ptashne, M. (1987). *A new class of yeast transcriptional activators*. Cell 51, 113-119.

659. Chen, J.-L. et al. (1994). *Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators*. Cell 79, 93-105.

1713. Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H., and Kornberg, R. D. (1994). *A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II.* Cell 77, 599-608.

2411. Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafsson, C. M., and Kornberg, R. D. (1999). *Conserved structures of mediator and RNA polymerase II holoenzyme.* Science 283, 985-987.

2412. Dotson, M. R., Yuan, C. X., Roeder, R. G., Myers, L. C., Gustafsson, C. M., Jiang, Y. W., Li, Y., Kornberg, R. D., and Asturias, F. J. (2000). *Structural organization of yeast and mammalian mediator complexes.* Proc. Natl. Acad. Sci. USA 97, 14307-14310.

ACTIVATING TRANSCRIPTION

5.22.6 Some promoter-binding proteins are repressors

Key Concepts

- Repression is usually achieved by affecting chromatin structure, but there are repressors that act by binding to specific promoters.

Repression of transcription in eukaryotes is generally accomplished at the level of influencing chromatin structure; regulator proteins that function like *trans*-acting bacterial repressors to block transcription are relatively rare, but some examples are known. One case is the global repressor NC2/Dr1/DRAP1, a heterodimer that binds to TBP to prevent it from interacting with other components of the basal apparatus (1741; 1742; 1743). The importance of this interaction is suggested by the lethality of null mutations in the genes that code for the repressor in yeast. Repressors that work in this way have an active role in inhibiting basal apparatus function.

In a more specific case, the CAAT sequence is a target for regulation. Two copies of this element are found in the promoter of a gene for histone H2B (see **Figure 21.22**) that is expressed only during spermatogenesis in a sea urchin. CAAT-binding factors can be extracted from testis tissue and also from embryonic tissues, but only the former can bind to the CAAT box. In the embryonic tissues, another protein, called the CAAT-displacement protein (CDP), binds to the CAAT boxes, *preventing the activator from recognizing them.*

Figure 22.10 illustrates the consequences for gene expression. In testis, the promoter is bound by transcription factors at the TATA box, CAAT boxes, and octamer sequences. In embryonic tissue, the exclusion of the CAAT-binding factor from the promoter prevents a transcription complex from being assembled. The analogy with the effect of a bacterial repressor in preventing RNA polymerase from initiating at the promoter is obvious. These results also make the point that the function of a protein in binding to a known promoter element cannot be assumed: it may be an activator, a repressor, or even irrelevant to gene transcription.

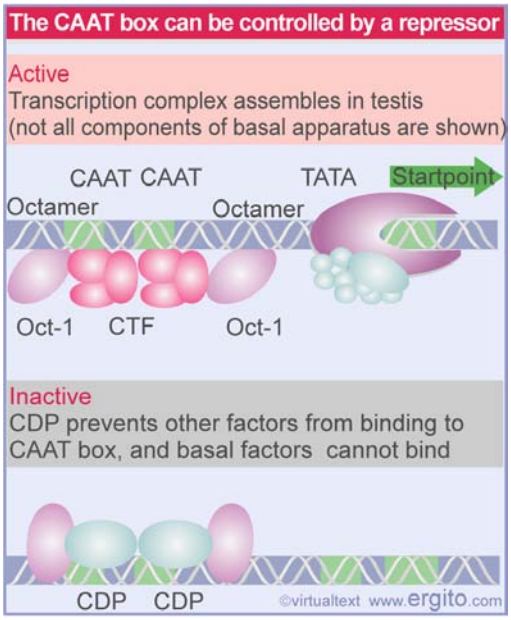

Figure 22.10 A transcription complex involves recognition of several elements in the sea urchin H2B promoter in testis. Binding of the CAAT displacement factor in embryo prevents the CAAT-binding factor from binding, so an active complex cannot form.

References

1741. Inostroza, J. A., Mermelstein, F. H., Ha, I., Lane, W. S., and Reinberg, D. (1992). *Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription.* Cell 70, 477-489.

1742. Goppelt, A., Stelzer, G., Lottspeich, F., and Meisterernst, M. (1996). *A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains.* EMBO J. 15, 3105-3116.

1743. Kim, T. K., Kim, T. K., Zhao, Y., Ge, H., Bernstein, R., and Roeder, R. G. (1995). *TATA-binding protein residues implicated in a functional interplay between negative cofactor NC2 (Dr1) and general factors TFIIA and TFIIB.* J. Biol. Chem. 270, 10976-10981.

ACTIVATING TRANSCRIPTION

5.22.7 Response elements are recognized by activators

Key Terms

A **response element** is a sequence in a eukaryotic promoter or enhancer that is recognized by a specific transcription factor.

The **heat shock response element (HSE)** is a sequence in a promoter or enhancer that is used to activate a gene by an activator induced by heat shock.

The **glucocorticoid response element (GRE)** is a sequence in a promoter or enhancer that is recognized by the glucocorticoid receptor ,which is activated by glucocorticoid steroids.

The **serum response element (SRE)** is a sequence in a promoter or enhancer that is activated by transcription factor(s) induced by treatment with serum. This activates genes that stimulate cell growth.

Heat shock genes are a set of loci that are activated in response to an increase in temperature (and other abuses to the cell). They occur in all organisms. They usually include chaperones that act on denatured proteins.

Key Concepts

- Response elements may be located in promoters or enhancers.
- Each response element is recognized by a specific activator.
- A promoter may have many response elements, which may activate transcription independently or in certain combinations.

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter (or enhancer) element that is recognized by an activator.* An element that causes a gene to respond to such a factor is called a **response element**; examples are the **HSE** (heat shock response element), **GRE** (glucocorticoid response element), **SRE** (serum response element). Response elements contain short consensus sequences; copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. Binding of an activator to the response element is required to allow RNA polymerase to initiate transcription. The difference from constitutively active activators is that the protein is either available

or is active only under certain conditions, which determine when the gene is to be expressed.

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the **heat shock genes**, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alternative –10 sequence common to the promoters of heat shock genes (see *Molecular Biology 3.9.16 Substitution of sigma factors may control initiation*). In eukaryotes, the heat shock genes also possess a common consensus sequence (HSE), but it is located at various positions relative to the startpoint, and is recognized by an independent activator, HSTF. The activation of this factor therefore provides a means to initiate transcription at the specific group of ~20 genes that contains the appropriate target sequence at its promoter.

All the heat shock genes of *D. melanogaster* contain multiple copies of the HSE. The HSTF binds cooperatively to adjacent response elements. Both the HSE and HSTF have been conserved in evolution, and it is striking that a heat shock gene from *D. melanogaster* can be activated in species as distant as mammals or sea urchins. The HSTF proteins of fruit fly and yeast appear similar, and show the same footprint pattern on DNA containing HSE sequences. Yeast HSTF becomes phosphorylated when cells are heat-shocked; this modification is responsible for activating the protein.

The metallothionein (MT) gene provides an example of how a single gene may be regulated by many different circuits. The metallothionein protein protects the cell against excess concentrations of heavy metals, by binding the metal and removing it from the cell. The gene is expressed at a basal level, but is induced to greater levels of expression by heavy metal ions (such as cadmium) or by glucocorticoids. The control region combines several different kinds of regulatory elements.

The organization of the promoter for a MT gene is summarized in **Figure 22.11**. A major feature of this map is the high density of elements that can activate transcription. The TATA and GC boxes are located at their usual positions fairly close to the startpoint. Also needed for the basal level of expression are the two basal level elements (BLE), which fit the formal description of enhancers. Although located near the startpoint, they can be moved elsewhere without loss of effect. They contain sequences related to those found in other enhancers, and are bound by proteins that bind the SV40 enhancer.

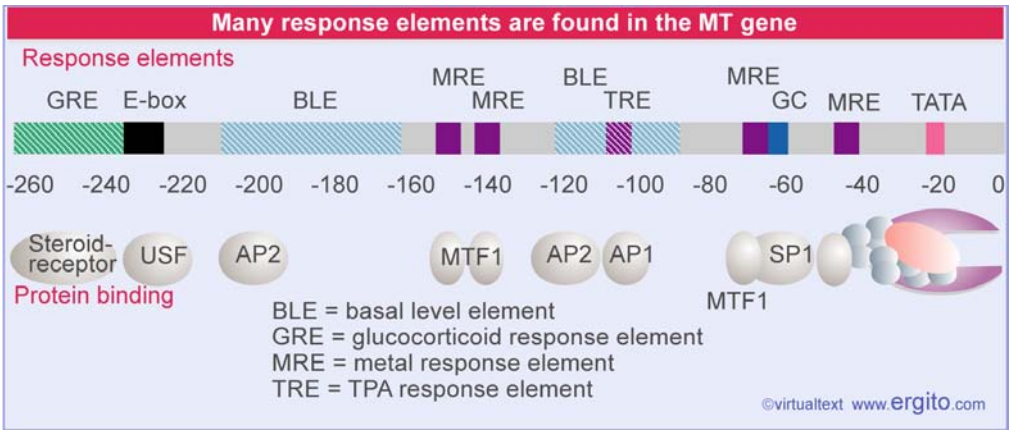

Figure 22.11 The regulatory region of a human metallothionein gene contains regulator elements in both its promoter and enhancer. The promoter has elements for metal induction; an enhancer has an element for response to glucocorticoid. Promoter elements are shown above the map, and proteins that bind them are indicated below.

The TRE is a consensus sequence that is present in several enhancers, including one BLE of metallothionein and the 72 bp repeats of the virus SV40. The TRE has a binding site for factor AP1; this interaction is part of the mechanism for constitutive expression, for which AP1 is an activator. However, AP1 binding also has a second function. The TRE confers a response to phorbol esters such as TPA (an agent that promotes tumors), and this response is mediated by the interaction of AP1 with the TRE. This binding reaction is one (not necessarily the sole) means by which phorbol esters trigger a series of transcriptional changes.

The inductive response to metals is conferred by the multiple MRE sequences, which function as promoter elements. The presence of one MRE confers the ability to respond to heavy metal; a greater level of induction is achieved by the inclusion of multiple elements. The factor MTF1 binds to the MRE in response to the presence of metal ions.

The response to steroid hormones is governed by a GRE, located 250 bp upstream of the startpoint, which behaves as an enhancer. Deletion of this region does not affect the basal level of expression or the level induced by metal ions. But it is absolutely needed for the response to steroids.

The regulation of metallothionein illustrates the general principle that *any one of several different elements, located in either an enhancer or promoter, can independently activate the gene.* The absence of an element needed for one mode of activation does not affect activation in other modes. The variety of elements, their independence of action, and the apparently unlimited flexibility of their relative arrangements, suggest that a factor binding to any one element is able independently to increase the efficiency of initiation by the basal transcription apparatus, probably by virtue of protein-protein interactions that stabilize or otherwise assist formation of the initiation complex.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.22.7

ACTIVATING TRANSCRIPTION

5.22.8 There are many types of DNA-binding domains

Key Terms

The **zinc finger** is a DNA-binding motif that typifies a class of transcription factor.

Steroid receptors are transcription factors that are activated by binding of a steroid ligand.

The **helix-turn-helix** motif describes an arrangement of two α helices that form a site that binds to DNA, one fitting into the major groove of DNA and other lying across it.

The **homeodomain** is a DNA-binding motif that typifies a class of transcription factors. The DNA sequence that codes for it is called the homeobox.

The **helix-loop-helix (HLH)** motif is responsible for dimerization of a class of transcription factors called HLH proteins. A bHLH protein has a basic DNA-binding sequence close to the dimerization motif.

The **leucine zipper** is a dimerization motif adjacent to a basic DNA-binding region that is found in a class of transcription factors.

Key Concepts

- Activators are classified according to the type of DNA-binding domain.
- Members of the same group have sequence variations of a specific motif that confer specificity for individual target sites.

It is common for an activator to have a modular structure in which different domains are responsible for binding to DNA and for activating transcription. Factors are often classified according to the type of DNA-binding domain (for review see 235; 238). Typically a relatively short motif in this domain is responsible for binding to DNA:

- The **zinc finger** motif comprises a DNA-binding domain. It was originally recognized in factor $TF_{III}A$, which is required for RNA polymerase III to transcribe 5S rRNA genes. It has since been identified in several other transcription factors (and presumed transcription factors). A distinct form of the motif is found also in the steroid receptors (672).
- The **steroid receptors** are defined as a group by a functional relationship: each receptor is activated by binding a particular steroid. The glucocorticoid receptor is the most fully analyzed. Together with other receptors, such as the thyroid hormone receptor or the retinoic acid receptor, the steroid receptors are members of the superfamily of ligand-activated activators with the same general *modus operandi: the protein factor is inactive until it binds a small ligand.*
- The **helix-turn-helix** motif was originally identified as the DNA-binding domain

of phage repressors. One α-helix lies in the major groove of DNA; the other lies at an angle across DNA. A related form of the motif is present in the **homeodomain**, a sequence first characterized in several proteins coded by genes concerned with developmental regulation in *Drosophila.* It is also present in genes for mammalian transcription factors.

- The amphipathic **helix-loop-helix** (**HLH**) motif has been identified in some developmental regulators and in genes coding for eukaryotic DNA-binding proteins. Each amphipathic helix presents a face of hydrophobic residues on one side and charged residues on the other side. The length of the connecting loop varies from 12-28 amino acids. The motif enables proteins to dimerize, and a basic region near this motif contacts DNA.

- **Leucine zippers** consist of a stretch of amino acids with a leucine residue in every seventh position. A leucine zipper in one polypeptide interacts with a zipper in another polypeptide to form a dimer. Adjacent to each zipper is a stretch of positively charged residues that is involved in binding to DNA.

The activity of an inducible activator may be regulated in any one of several ways, as illustrated schematically in **Figure 22.12**:

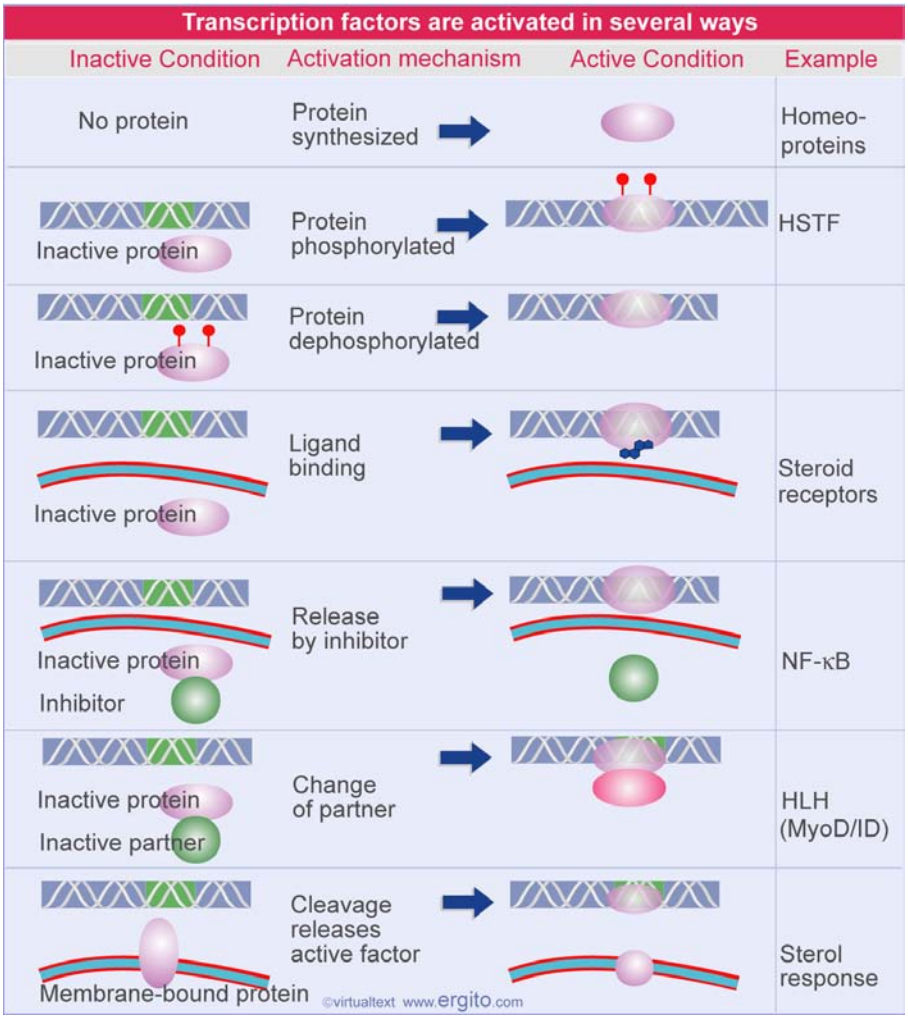

Figure 22.12 The activity of a regulatory transcription factor may be controlled by synthesis of protein, covalent modification of protein, ligand binding, or binding of inhibitors that sequester the protein or affect its ability to bind to DNA.

- A factor is tissue-specific because it is synthesized only in a particular type of cell. This is typical of factors that regulate development, such as homeodomain proteins.
- The activity of a factor may be directly controlled by modification. HSTF is converted to the active form by phosphorylation. AP1 (a heterodimer between the subunits Jun and Fos) is converted to the active form by phosphorylating the Jun subunit.
- A factor is activated or inactivated by binding a ligand. The steroid receptors are prime examples. Ligand binding may influence the localization of the protein (causing transport from cytoplasm to nucleus), as well as determining its ability to bind to DNA.
- Availability of a factor may vary; for example, the factor NF- κB (which activates immunoglobulin κ genes in B lymphocytes) is present in many cell

types. But it is sequestered in the cytoplasm by the inhibitory protein I- κB. In B lymphocytes, NF- κB is released from I- κB and moves to the nucleus, where it activates transcription.

- An extreme example of control of availability is found when a factor is actually part of a cytoplasmic structure, and is released from that structure to translocate to the nucleus.

- A dimeric factor may have alternative partners. One partner may cause it to be inactive; synthesis of the active partner may displace the inactive partner. Such situations may be amplified into networks in which various alternative partners pair with one another, especially among the HLH proteins (675).

- The factor may be cleaved from an inactive precursor. One activator is produced as a protein bound to the nuclear envelope and endoplasmic reticulum. The absence of sterols (such as cholesterol) causes the cytosolic domain to be cleaved; it then translocates to the nucleus and provides the active form of the activator.

(We note *en passant* that mutations of the transcription factors in some of these classes give rise to factors that inappropriately activate, or prevent activation, of transcription; their roles in generating tumors are discussed in *Molecular Biology 6.30.18 Oncoproteins may regulate gene expression*, and **Figure 30.26** should be compared with **Figure 22.12**.)

We now discuss in more detail the DNA-binding and activation reactions that are sponsored by some of these classes of proteins.

Reviews

235. Harrison, S. C. (1991). *A structural taxonomy of DNA-binding proteins*. Nature 353, 715-719.

238. Pabo, C. T. and Sauer, R. T. (1992). *Transcription factors: structural families and principles of DNA recognition*. Annu. Rev. Biochem. 61, 1053-1095.

References

672. Miller, J. et al. (1985). *Repetitive zinc binding domains in the protein transcription factor IIIA from Xenopus oocytes*. EMBO J. 4, 1609-1614.

675. Murre, C., McCaw, P. S., and Baltimore, D. (1989). *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins*. Cell 56, 777-783.

ACTIVATING TRANSCRIPTION

5.22.9 A zinc finger motif is a DNA-binding domain

Key Concepts

- A zinc finger is a loop of ~23 amino acids that protrudes from a zinc-binding site formed by His and Cys amino acids.
- A zinc finger protein usually has multiple zinc fingers.
- The C-terminal part of each finger forms an α-helix that bind one turn of the major groove of DNA.
- Some zinc finger proteins bind RNA instead of or as well as DNA.

Zinc fingers take their name from the structure illustrated in **Figure 22.13**, in which a small group of conserved amino acids binds a zinc ion to form an independent domain in the protein. Two types of DNA-binding proteins have structures of this type: the classic "zinc finger" proteins; and the steroid receptors.

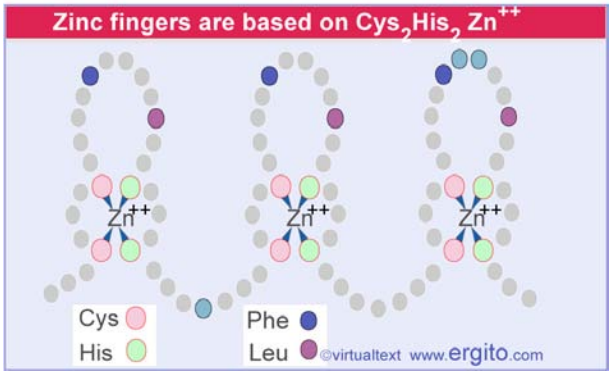

Figure 22.13 Transcription factor SP1 has a series of three zinc fingers, each with a characteristic pattern of cysteine and histidine residues that constitute the zinc-binding site.

A "finger protein" typically has a series of zinc fingers, as depicted in the figure. The consensus sequence of a single finger is:

Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His

The motif takes its name from the loop of amino acids that protrudes from the zinc-binding site and is described as the Cys_2/His_2 finger. The zinc is held in a tetrahedral structure formed by the conserved Cys and His residues. The finger itself comprises ~23 amino acids, and the linker between fingers is usually 7-8 amino acids.

Zinc fingers are a common motif in DNA-binding proteins. The fingers usually are

organized as a single series of tandem repeats; occasionally there is more than one group of fingers. The stretch of fingers ranges from 9 repeats that occupy almost the entire protein (as in $TF_{III}A$) to providing just one small domain consisting of 2 fingers (as in the *Drosophila* regulator ADR1). The activator Sp1 has a DNA-binding domain that consists of 3 zinc fingers (656; 672).

The crystal structure of DNA bound by a protein with three fingers suggests the structure illustrated schematically in **Figure 22.14**. The C-terminal part of each finger forms α-helices that bind DNA; the N-terminal part forms a β-sheet. (For simplicity, the β-sheet and the location of the zinc ion are not shown in the lower part of the figure.) The three α-helical stretches fit into one turn of the major groove; each α-helix (and thus each finger) makes two sequence-specific contacts with DNA (indicated by the arrows). We expect that the nonconserved amino acids in the C-terminal side of each finger are responsible for recognizing specific target sites (677).

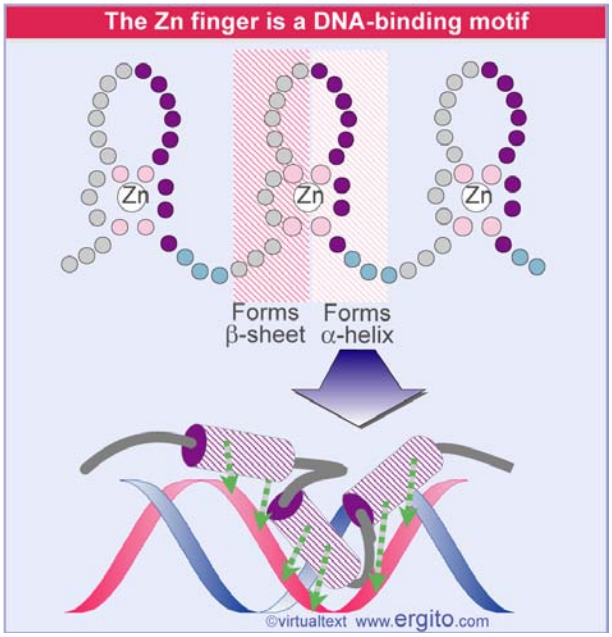

Figure 22.14 Zinc fingers may form α -helices that insert into the major groove, associated with β -sheets on the other side.

Knowing that zinc fingers are found in authentic activators that assist both RNA polymerases II and III, we may view finger proteins from the reverse perspective. When a protein is found to have multiple zinc fingers, there is at least a *prima facie* case for investigating a possible role as a transcription factor. Such an identification has suggested that several loci involved in embryonic development of *D. melanogaster* are regulators of transcription.

However, it is necessary to be cautious about interpreting the presence of (putative) zinc fingers, especially when the protein contains only a single finger motif. Fingers may be involved in binding RNA rather than DNA or even unconnected with any nucleic acid binding activity. For example, the prototype zinc finger protein, $TF_{III}A$, binds both to the 5S gene and to the product, 5S rRNA. A translation initiation

factor, eIF2 β has a zinc finger; and mutations in the finger influence the recognition of initiation codons. Retroviral capsid proteins have a motif related to the finger that may be involved in binding the viral RNA.

References

656. Kadonaga, J. et al. (1987). *Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain*. Cell 51, 1079-1090.

672. Miller, J. et al. (1985). *Repetitive zinc binding domains in the protein transcription factor IIIA from Xenopus oocytes*. EMBO J. 4, 1609-1614.

677. Pavletich, N. P. and Pabo, C. O. (1991). *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 21* Å. Science 252, 809-817.

ACTIVATING TRANSCRIPTION

5.22.10 Steroid receptors are activators

Key Concepts

- Steroid receptors are examples of ligand-responsive activators that are activated by binding a steroid (or other related molecules).
- There are separate DNA-binding and ligand-binding domains.

Steroid hormones are synthesized in response to a variety of neuroendocrine activities, and exert major effects on growth, tissue development, and body homeostasis in the animal world. The major groups of steroids and some other compounds with related (molecular) activities are classified in **Figure 22.15**.

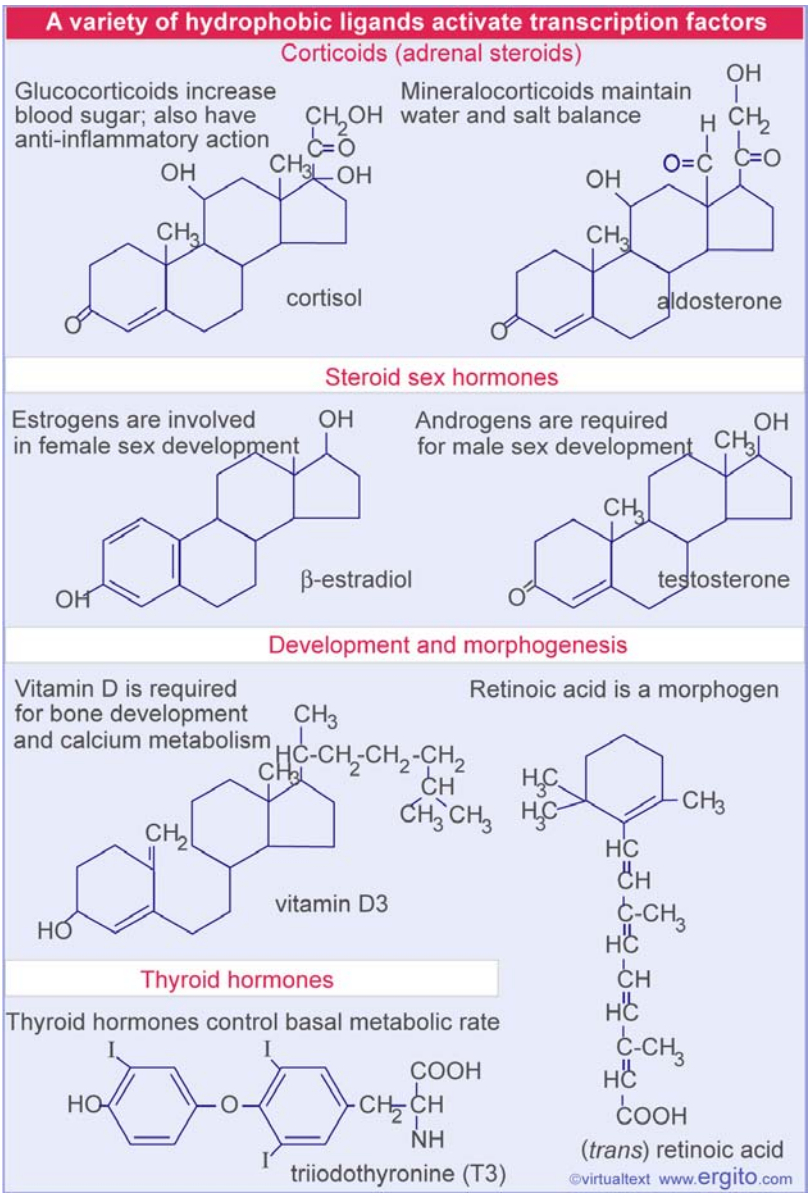

Figure 22.15 Several types of hydrophobic small molecules activate transcription factors.

The adrenal gland secretes >30 steroids, the two major groups being the glucocorticoids and mineralocorticoids. Steroids provide the reproductive hormones (androgen male sex hormones and estrogen female sex hormones). Vitamin D is required for bone development.

Other hormones, with unrelated structures and physiological purposes, function at the molecular level in a similar way to the steroid hormones. Thyroid hormones, based on iodinated forms of tyrosine, control basal metabolic rate in animals. Steroid and thyroid hormones also may be important in metamorphosis (ecdysteroids in insects, and thyroid hormones in frogs).

Retinoic acid (vitamin A) is a morphogen responsible for development of the anterior-posterior axis in the developing chick limb bud. Its metabolite, 9-*cis* retinoic acid, is found in tissues that are major sites for storage and metabolism of vitamin A.

We may account for these various actions in terms of pathways for regulating gene expression. These diverse compounds share a common mode of action: *each is a small molecule that binds to a specific receptor that activates gene transcription.* ("Receptor" may be a misnomer: the protein is a receptor for steroid or thyroid hormone in the same sense that *lac* repressor is a receptor for a β-galactoside: it is not a receptor in the sense of comprising a membrane-bound protein that is exposed to the cell surface.)

Receptors for the diverse groups of steroid hormones, thyroid hormones, and retinoic acid represent a new "superfamily" of gene regulators, the ligand-responsive activators. All the receptors have independent domains for DNA-binding and hormone binding, in the same relative locations. Their general organization is summarized in **Figure 22.16** (for review see 231).

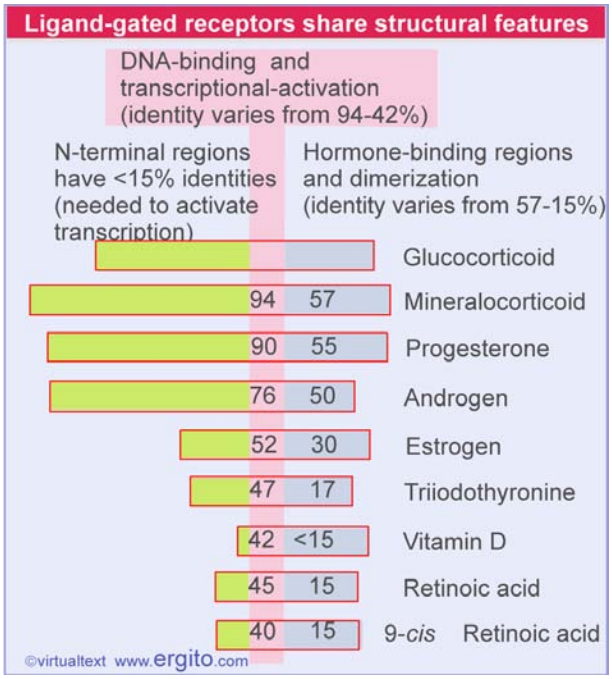

Figure 22.16 Receptors for many steroid and thyroid hormones have a similar organization, with an individual N-terminal region, conserved DNA-binding region, and a C-terminal hormone-binding region.

The central part of the protein is the DNA-binding domain. These regions are closely related for the various steroid receptors (from the most closely related pair with 94% sequence identity to the least well related pair at 42% identity). The act of binding DNA cannot be disconnected from the ability to activate transcription, because mutations in this domain affect both activities.

The N-terminal regions of the receptors show the least conservation of sequence. They include other regions that are needed to activate transcription.

The C-terminal domains bind the hormones. Those in the steroid receptor family show identities ranging from 30-57%, reflecting specificity for individual hormones. Their relationships with the other receptors are minimal, reflecting specificity for a variety of compounds – thyroid hormones, vitamin D, retinoic acid, etc. This domain also has the motifs responsible for dimerization and a region involved in transcriptional activation (for review see 1436).

Some ligands have multiple receptors that are closely related, such as the 3 retinoic acid receptors (RAR α , β, γ) and the three receptors for 9-*cis*-retinoic acid (RXR α , β, γ).

Last updated on 2-12-2001

Reviews

231. Evans, R. M. (1988). *The steroid and thyroid hormone receptor superfamily*. Science 240, 889-895.

1436. Mangelsdorf, D. J. and Evans, R. (1995). *The RXR heterodimers and orphan receptors.* Cell 83, 841-850.

ACTIVATING TRANSCRIPTION

5.22.11 Steroid receptors have zinc fingers

Key Concepts

- The DNA binding domain of a steroid receptor is a type of zinc finger that has Cys but not His residues.
- Glucocorticoid and estrogen receptors each have two zinc fingers, the first of which determines the DNA target sequence.
- Steroid receptors bind to DNA as dimers.

Steroid receptors (and some other proteins) have another type of zinc finger that is different from Cys_2/His_2 fingers. The structure is based on a sequence with the zinc-binding consensus:

$Cys\text{-}X_2\text{-}Cys\text{-}X_{13}\text{-}Cys\text{-}X_2\text{-}Cys$

These are called Cys_2/Cys_2 fingers. Proteins with Cys_2/Cys_2 fingers often have nonrepetitive fingers, in contrast with the tandem repetition of the Cys_2/His_2 type. Binding sites in DNA (where known) are short and palindromic.

The glucocorticoid and estrogen receptors each have two fingers, each with a zinc atom at the center of a tetrahedron of cysteines. The two fingers form α-helices that fold together to form a large globular domain. The aromatic sides of the α-helices form a hydrophobic center together with a β-sheet that connects the two helices. One side of the N-terminal helix makes contacts in the major groove of DNA. Two glucocorticoid receptors dimerize upon binding to DNA, and each engages a successive turn of the major groove. This fits with the palindromic nature of the response element (see *Molecular Biology 5.22.13 Steroid receptors recognize response elements by a combinatorial code*).

Each finger controls one important property of the receptor. **Figure 22.18** identifies the relevant amino acids. Those on the right side of the first finger determine the sequence of the target in DNA; those on the left side of the second finger control the spacing between the target sites recognized by each subunit in the dimer (see *Molecular Biology 5.22.13 Steroid receptors recognize response elements by a combinatorial code*).

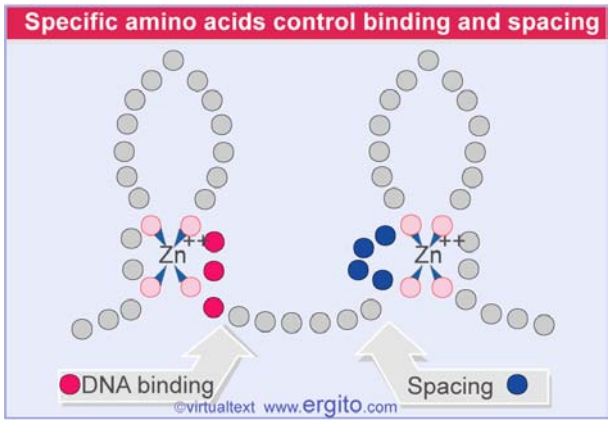

Figure 22.18 The first finger of a steroid receptor controls which DNA sequence is bound (positions shown in red); the second finger controls spacing between the sequences (positions shown in blue).

Direct evidence that the first finger binds DNA was obtained by a "specificity swap" experiment. The finger of the estrogen receptor was deleted and replaced by the sequence of the glucocorticoid receptor. The new protein recognized the GRE sequence (the usual target of the glucocorticoid receptor) instead of the ERE (the usual target of the estrogen receptor). This region therefore establishes the specificity with which DNA is recognized.

The differences between the sequences of the glucocorticoid receptor and estrogen receptor fingers lie mostly at the base of the finger. The substitution at two positions shown in **Figure 22.17** allows the glucocorticoid receptor to bind at an ERE instead of a GRE (1768; for review see 240).

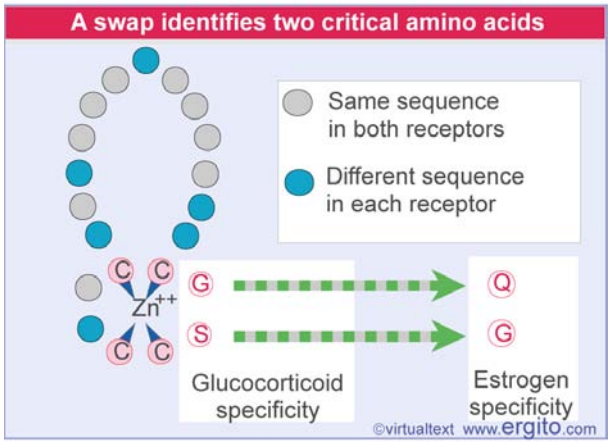

Figure 22.17 Discrimination between GRE and ERE target sequences is determined by two amino acids at the base of the first zinc finger in the receptor.

Reviews

240. Tsai, M J, and O'Malley, B W (1994). *Molecular mechanisms of action of steroid/thyroid receptor superfamily members.* Annu. Rev. Biochem. 63, 451-486.

References

1768. Umesono, K. and Evans, R. M. (1989). *Determinants of target gene specificity for steroid/thyroid hormone receptors.* Cell 57, 1139-1146.

ACTIVATING TRANSCRIPTION

5.22.12 Binding to the response element is activated by ligand-binding

Key Concepts

- Binding of ligand to the C-terminal domain increases the affinity of the DNA-binding domain for its specific target site in DNA.

We know most about the interaction of glucocorticoids with their receptor, whose action is illustrated in **Figure 22.19**. A steroid hormone can pass through the cell membrane to enter the cell by simple diffusion. Within the cell, a glucocorticoid binds the glucocorticoid receptor. (Work on the glucocorticoid receptor has relied on the synthetic steroid hormone, dexamethasone.) The localization of free receptors is not entirely clear; they may be in equilibrium between the nucleus and cytoplasm. But when hormone binds to the receptor, the protein is converted into an activated form that has an increased affinity for DNA, so the hormone-receptor complex is always localized in the nucleus.

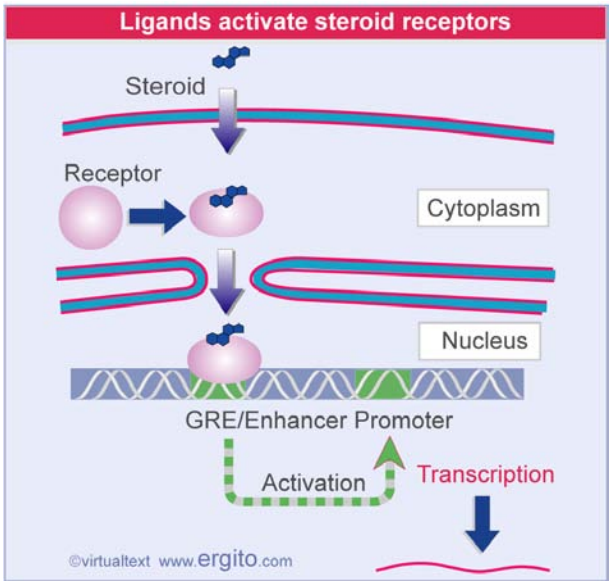

Figure 22.19 Glucocorticoids regulate gene transcription by causing their receptor to bind to an enhancer whose action is needed for promoter function.

The activated receptor recognizes a specific consensus sequence that identifies the GRE, the glucocorticoid response element. The GRE is typically located in an enhancer that may be several kb upstream or downstream of the promoter. When the steroid-receptor complex binds to the enhancer, the nearby promoter is activated, and transcription initiates there. Enhancer activation provides the general mechanism by which steroids regulate a wide set of target genes.

The C-terminal region regulates the activity of the receptor in a way that varies for the individual receptor. If the C-terminal domain of the glucocorticoid receptor is deleted, the remaining N-terminal protein is constitutively active: it no longer requires steroids for activity. This suggests that, in the absence of steroid, the steroid-binding domain prevents the receptor from recognizing the GRE; it functions as an internal negative regulator. The addition of steroid inactivates the inhibition, releasing the receptor's ability to bind the GRE and activate transcription. The basis for the repression could be internal, relying on interactions with another part of the receptor. Or it could result from an interaction with some other protein, which is displaced when steroid binds.

The interaction between the domains is different in the estrogen receptor. If the hormone-binding domain is deleted, the protein is unable to activate transcription, although it continues to bind to the ERE. This region is therefore required to activate rather than to repress activity.

ACTIVATING TRANSCRIPTION

5.22.13 Steroid receptors recognize response elements by a combinatorial code

Key Concepts

- A steroid response element consists of two short half sites that may be palindromic or directly repeated.
- There are only two types of half sites.
- A receptor recognizes its response element by the orientation and spacing of the half sites.
- The sequence of the half site is recognized by the first zinc finger.
- The second zinc finger is responsible for dimerization, which determines the distance between the subunits.
- Subunit separation in the receptor determines the recognition of spacing in the response element.
- Some steroid receptors function as homodimers but others form heterodimers.
- Homodimers recognize palindromic response elements; heterodimers recognize response elements with directly repeated half sites.

Each receptor recognizes a response element that consists of two short repeats (or half sites). This immediately suggests that the receptor binds as a dimer, so that each half of the consensus is contacted by one subunit (reminiscent of the λ operator-repressor interaction described in *Molecular Biology 3.12.12 Repressor uses a helix-turn-helix motif to bind DNA*).

The half sites may be arranged either as palindromes or as repeats in the same orientation. They are separated by 0-4 base pairs whose sequence is irrelevant. Only two types of half site are used by the various receptors. Their orientation and spacing determine which receptor recognizes the response element. This behavior allows response elements that have restricted consensus sequences to be recognized specifically by a variety of receptors. The rules that govern recognition are not absolute, but may be modified by context, and there are also cases in which palindromic response elements are recognized permissively by more than one receptor (for review see 229).

The receptors fall into two groups:

- Glucocorticoid (GR), mineralocorticoid (MR), androgen (AR), and progesterone (PR) receptors all form homodimers. They recognize response elements whose half sites have the consensus sequence TGTTCT. **Figure 22.20** shows that the

half sites are arranged as palindromes, and the spacing between the sites determines the type of element. The estrogen (ER) receptor functions in the same way, but has the half site sequence TGACCT.

- The 9-*cis*-retinoic acid (RXR) receptor forms homodimers and also forms heterodimers with ~15 other receptors, including thyroid (T3R), vitamin D (VDR), and retinoic acid (RAR) (1436). **Figure 22.21** shows that the dimers recognize half elements with the sequence TGACCT. The half sites are arranged as direct repeats, and recognition is controlled by spacing between them (1769). Some of the heterodimeric receptors are activated when the ligand binds to the partner for RXR; others can be activated by ligand binding either to this subunit or to the RXR subunit. These receptors can also form homodimers, which recognize palindromic sequences.

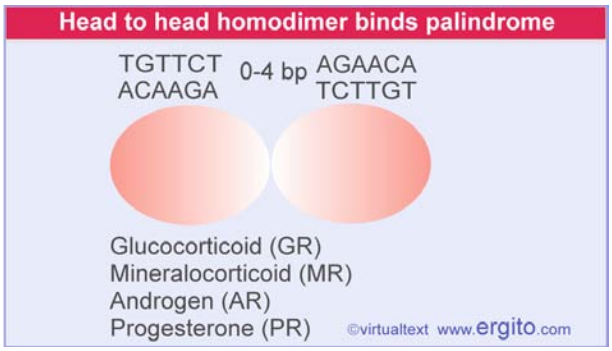

Figure 22.20 Response elements formed from the palindromic half site TGTTCT are recognized by several different receptors depending on the spacing between the half sites.

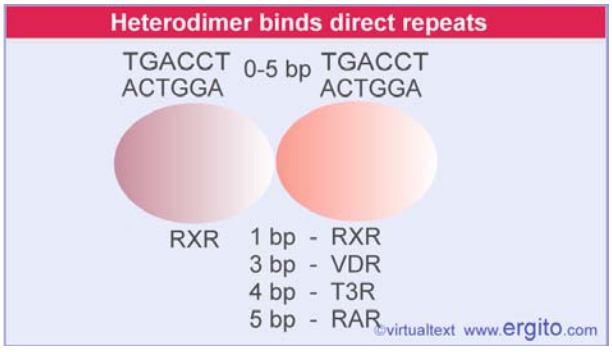

Figure 22.21 Response elements with the direct repeat TGACCT are recognized by heterodimers of which one member is RXR.

Now we are in a position to understand the basis for specificity of recognition. Recall that **Figure 22.18** shows how recognition of the sequence of the half site is conferred by the amino acid sequence in the first finger. Specificity for the spacing between half sites is carried by amino acids in the second finger. The structure of the dimer determines the distance between the subunits that sit in successive turns of the major groove, and thus controls the response to the spacing of half sites (679). The exact positions of the residues responsible for dimerization differ in individual pairwise combinations.

How do the steroid receptors activate transcription? They do not act directly on the basal apparatus, but function via a coactivating complex. The coactivator includes various activities, including the common component CBP/p300, one of whose functions is to modify the structure of chromatin by acetylating histones (see **Figure 23.13**).

All receptors in the superfamily are ligand-dependent activators of transcription. However, some are also able to repress transcription. The TR and RAR receptors, in the form of heterodimers with RXR, bind to certain loci in the *absence* of ligand and repress transcription by means of their ability to interact with a corepressor protein. The corepressor functions by the reverse of the mechanism used by coactivators: it inhibits the function of the basal transcription apparatus, one of its actions being the deacetylation of histones (see **Figure 23.15**). We do not know the relative importance of the repressor activity *vis*-À-*vis* the ligand-dependent activation in the physiological response to hormone (660).

The effect of ligand binding on the receptor is to convert it from a repressing complex to an activating complex, as shown in **Figure 22.22**. In the absence of ligand, the receptor is bound to a corepressor complex. The component of the corepressor that binds to the receptor is SMRT. Binding of ligand causes a conformational change that displaces SMRT. This allows the coactivator to bind.

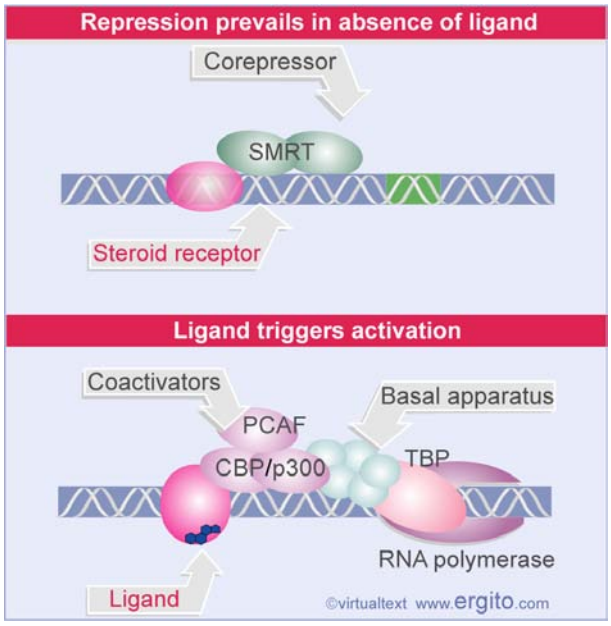

Figure 22.22 TR and RAR bind the SMRT corepressor in the absence of ligand. The promoter is not expressed. When SMRT is displaced by binding of ligand, the receptor binds a coactivator complex. This leads to activation of transcription by the basal apparatus.

Reviews

229. Yamamoto, K. R. (1985). *Steroid receptor regulated transcription of specific genes and gene networks*. Annu. Rev. Genet. 19, 209-252.

1436. Mangelsdorf, D. J. and Evans, R. (1995). *The RXR heterodimers and orphan receptors.* Cell 83, 841-850.

References

660. Hurlein, A. J. et al. (1995). *Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor corepressor*. Nature 377, 397-404.

679. Rastinejad, F., Perlmann, T., Evans, R. M., and Sigler, P. B. (1995). *Structural determinants of nuclear receptor assembly on DNA direct repeats*. Nature 375, 203-211.

1769. Umesono, K., Murakami, K. K., Thompson, C. C., and Evans, R. M. (1991). *Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors.* Cell 65, 1255-1266.

ACTIVATING TRANSCRIPTION

5.22.14 Homeodomains bind related targets in DNA

Key Concepts

- The homeodomain is a DNA-binding domain of 60 amino acids that has three α-helices.
- The C-terminal α-helix-3 is 17 amino acids and binds in the major groove of DNA.
- The N-terminal arm of the homeodomain projects into the minor groove of DNA.
- Proteins containing homeodomains may be either activators or repressors of transcription.

The homeobox is a sequence that codes for a domain of 60 amino acids present in proteins of many or even all eukaryotes. Its name derives from its original identification in *Drosophila* homeotic loci (whose genes determine the identity of body structures). It is present in many of the genes that regulate early development in *Drosophila*, and a related motif is found in genes in a wide range of higher eukaryotes. The homeodomain is found in many genes concerned with developmental regulation (see *Molecular Biology 6.31.22 The homeobox is a common coding motif in homeotic genes*). Sequences related to the homeodomain are found in several types of animal transcription factors.

In *Drosophila* homeotic genes, the homeodomain often (but not always) occurs close to the C-terminal end. Some examples of genes containing homeoboxes are summarized in **Figure 22.23**. Often the genes have little conservation of sequence except in the homeobox. The conservation of the homeobox sequence varies. A major group of homeobox-containing genes in *Drosophila* has a well conserved sequence, with 80-90% similarity in pairwise comparisons. Other genes have less closely related homeoboxes. The homeodomain is sometimes combined with other motifs in animal transcription factors. One example is presented by the Oct (octamer-binding) proteins, in which a conserved stretch of 75 amino acids called the Pou region is located close to a region resembling the homeodomain. The homeoboxes of the Pou group of proteins are the least closely related to the original group, and thus comprise the farthest extension of the family.

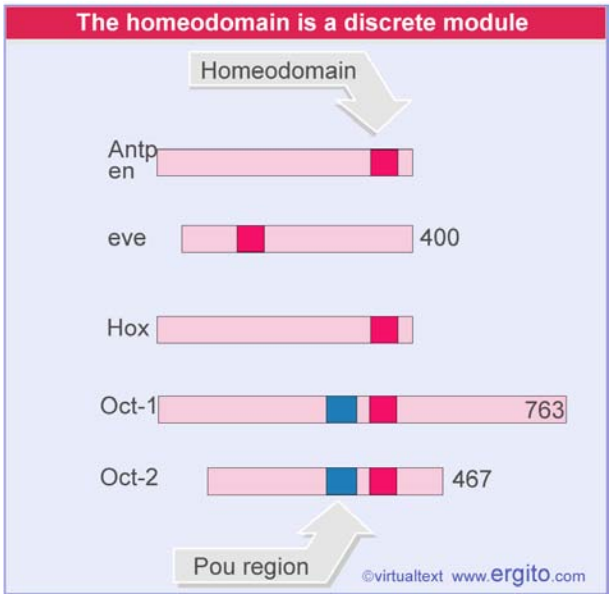

Figure 22.23 The homeodomain may be the sole DNA-binding motif in a transcriptional regulator or may be combined with other motifs. It represents a discrete (60 residue) part of the protein.

The homeodomain is responsible for binding to DNA, and experiments to swap homeodomains between proteins suggest that the specificity of DNA recognition lies within the homeodomain, but (like the situation with phage repressors) no simple code relating protein and DNA sequences can be deduced. The C-terminal region of the homeodomain shows homology with the helix-turn-helix motif of prokaryotic repressors. We recall from *Molecular Biology 3.12.12 Repressor uses a helix-turn-helix motif to bind DNA* that the λ repressor has a "recognition helix" (α-helix-3) that makes contacts in the major groove of DNA, while the other helix (α-helix-2) lies at an angle across the DNA. The homeodomain can be organized into three potential helical regions; the sequences of three examples are compared in **Figure 22.24**. The best conserved part of the sequence lies in the third helix. The difference between these structures and the prokaryotic repressor structures lies in the length of the helix that recognizes DNA, helix-3, which is 17 amino acids long in the homeodomain, compared to 9 residues long in the λ repressor.

Figure 22.24 The homeodomain of the *Antennapedia* gene represents the major group of genes containing homeoboxes in Drosophila; *engrailed* (*en*) represents another type of homeotic gene; and the mammalian factor Oct-2 represents a distantly related group of transcription factors. The homeodomain is conventionally numbered from 1 to 60. It starts with the N-terminal arm, and the three helical regions occupy residues 10-22, 28-38, and 42-58. Amino acids in red are conserved in all three examples.

The structure of the homeodomain of the *D. melanogaster* Engrailed protein is represented schematically in **Figure 22.25**. Helix 3 binds in the major groove of DNA and makes the majority of the contacts between protein and nucleic acid. Many of the contacts that orient the helix in the major groove are made with the phosphate backbone, so they are not specific for DNA sequence. They lie largely on one face of the double helix, and flank the bases with which specific contacts are made. The remaining contacts are made by the N-terminal arm of the homeodomain, the sequence that just precedes the first helix. It projects into the minor groove. So the N-terminal and C-terminal regions of the homeodomain are primarily responsible for contacting DNA (678).

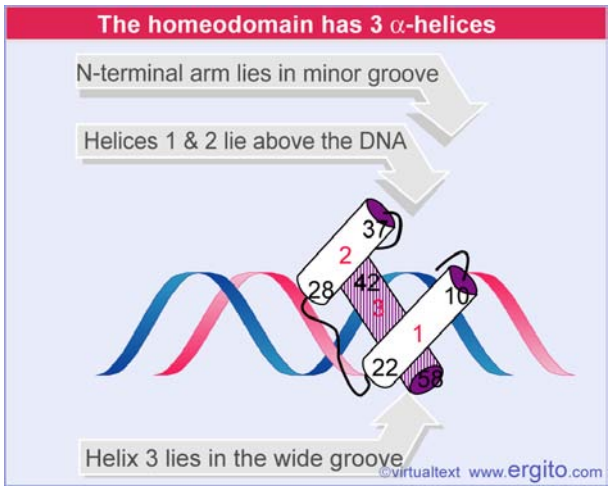

Figure 22.25 Helix 3 of the homeodomain binds in the major groove of DNA, with helices 1 and 2 lying outside the double helix. Helix 3 contacts both the phosphate backbone and specific bases. The N-terminal arm lies in the minor groove, and makes additional contacts.

A striking demonstration of the generality of this model derives from a comparison of the crystal structure of the homeodomain of engrailed with that of the α2 mating protein of yeast. The DNA-binding domain of this protein resembles a homeodomain, and can form three similar helices: its structure in the DNA groove can be superimposed almost exactly on that of the engrailed homeodomain. These similarities suggest that all homeodomains bind to DNA in the same manner. This means that a relatively small number of residues in helix-3 and in the N-terminal arm are responsible for specificity of contacts with DNA (for review see 239).

One group of homeodomain-containing proteins is the set of Hox proteins (see **Figure 31.39**). They bind to DNA with rather low sequence specificity, and it has been puzzling how these proteins can have different specificities. It turns out that Hox proteins often bind to DNA as heterodimers with a partner (called Exd in flies and Pbx in vertebrates). The heterodimer has a more restricted specificity *in vitro* than an individual Hox protein; typically it binds the 10 bp sequence TGATNNATNN. Still this is not enough to account for the differences in the specificities of Hox proteins. A third protein, Hth, which is necessary to localize Exd in the nucleus, also forms part of the complex that binds DNA, and may restrict the binding sites further. But since the same partners (Exd and Hth) are present together with each Hox protein in the trimeric complex, it remains puzzling how each Hox protein has sufficient specificity.

Homeodomain proteins can be either transcriptional activators or repressors. The nature of the factor depends on the other domain(s) – the homeodomain is responsible solely for binding to DNA. The activator or repressor domains both act by influencing the basal apparatus. Activator domains may interact with coactivators that in turn bind to components of the basal apparatus. Repressor domains also interact with the transcription apparatus (that is, they do not act by blocking access to DNA as such). The repressor Eve, for example, interacts directly with $TF_{II}D$ (674).

Reviews

239. Gehring, W. J. et al. (1994). *Homeodomain-DNA recognition*. Cell 78, 211-223.

References

674. Han, K., Levine, M. S., and Manley, J. L. (1989). *Synergistic activation and repression of transcription by Drosophila homeobox proteins*. Cell 56, 573-583.

678. Wolberger, C. et al. (1991). *Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions*. Cell 67, 517-528.

ACTIVATING TRANSCRIPTION

5.22.15 Helix-loop-helix proteins interact by combinatorial association

Key Terms

The **helix-loop-helix (HLH)** motif is responsible for dimerization of a class of transcription factors called HLH proteins. A bHLH protein has a basic DNA-binding sequence close to the dimerization motif.

A **bHLH protein** has a basic DNA-binding region adjacent to the helix-loop-helix motif.

Key Concepts

- Helix-loop-helix proteins have a motif of 40-50 amino acids that comprises two amphipathic α-helices of 15-16 residues separated by a loop.
- The helices are responsible for dimer formation.
- bHLH proteins have a basic sequence adjacent to the HLH motif that is responsible for binding to DNA.
- Class A bHLH proteins are ubiquitously expressed. Class B bHLH proteins are tissue specific.
- A class B protein usually forms a heterodimer with a class A protein.
- HLH proteins that lack the basic region prevent a bHLH partner in a heterodimer from binding to DNA.
- HLH proteins form combinatorial associations that may be changed during development by the addition or removal of specific proteins.

Two common features in DNA-binding proteins are the presence of helical regions that bind DNA, and the ability of the protein to dimerize. Both features are represented in the group of **helix-loop-helix** proteins that share a common type of sequence motif: a stretch of 40-50 amino acids contains two amphipathic α-helices separated by a linker region (the loop) of varying length. (An amphipathic helix forms two faces, one presenting hydrophobic amino acids, the other presenting charged amino acids.) The proteins in this group form both homodimers and heterodimers by means of interactions between the hydrophobic residues on the corresponding faces of the two helices (675). The helical regions are 15-16 amino acids long, and each contains several conserved residues. Two examples are compared in **Figure 22.26**. The ability to form dimers resides with these amphipathic helices, and is common to all HLH proteins. The loop is probably important only for allowing the freedom for the two helical regions to interact independently of one another.

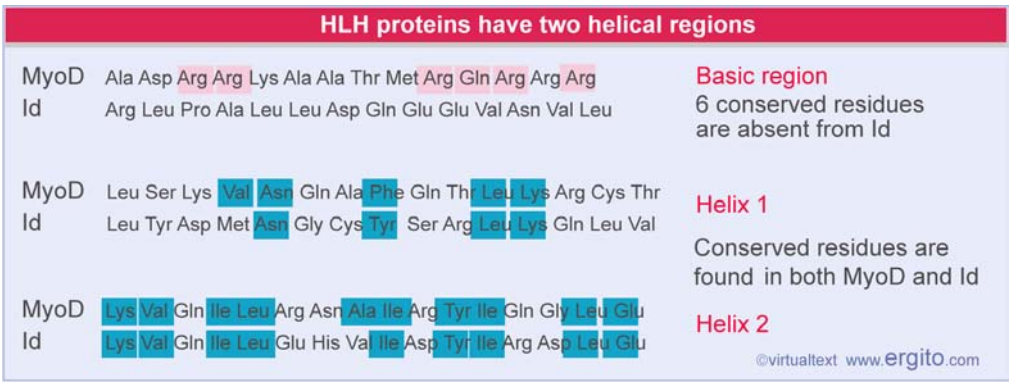

Figure 22.26 All HLH proteins have regions corresponding to helix 1 and helix 2, separated by a loop of 10-24 residues. Basic HLH proteins have a region with conserved positive charges immediately adjacent to helix 1.

Most HLH proteins contain a region adjacent to the HLH motif itself that is highly basic, and which is needed for binding to DNA. There are ~6 conserved residues in a stretch of 15 amino acids (see **Figure 22.26**). Members of the group with such a region are called **bHLH proteins**. A dimer in which both subunits have the basic region can bind to DNA. The HLH domains probably correctly orient the two basic regions contributed by the individual subunits.

The bHLH proteins fall into two general groups. Class A consists of proteins that are ubiquitously expressed, including mammalian E12/E47. Class B consists of proteins that are expressed in a tissue-specific manner, including mammalian MyoD, myogenin, and Myf-5 (a group of activators that are involved in myogenesis [muscle formation]). A common *modus operandi* for a tissue-specific bHLH protein is to form a heterodimer with a ubiquitous partner. There is also a group of gene products that specify development of the nervous system in *D. melanogaster* (where *Ac-S* is the tissue-specific component and *da* is the ubiquitous component). The Myc proteins (which are the cellular counterparts of oncogene products and are involved in growth regulation) form a separate class of bHLH proteins, whose partners and targets are different.

Dimers formed from bHLH proteins differ in their abilities to bind to DNA. For example, E47 homodimers, E12-E47 heterodimers, and MyoD-E47 heterodimers all form efficiently and bind strongly to DNA; E12 homodimerizes well but binds DNA poorly, while MyoD homodimerizes only poorly. So both dimer formation and DNA binding may represent important regulatory points. At this juncture, it is possible to define groups of HLH proteins whose members form various pairwise combinations, but not to predict from the sequences the strengths of dimer formation or DNA binding. All of the dimers in this group that bind DNA recognize the same consensus sequence, but we do not know yet whether different homodimers and heterodimers have preferences for slightly different target sites that are related to their functions.

Differences in DNA-binding result from properties of the region in or close to the HLH motif; for example, E12 differs from E47 in possessing an inhibitory region just by the basic region, which prevents DNA binding by homodimers. Some HLH proteins lack the basic region and/or contain proline residues that appear to disrupt its function. The example of the protein Id is shown in **Figure 22.26**. Proteins of this type have the same capacity to dimerize as bHLH proteins, but a dimer that contains

one subunit of this type can no longer bind to DNA specifically. This is a forceful demonstration of the importance of doubling the DNA-binding motif in DNA-binding proteins (670; 671; 675).

The importance of the distinction between the nonbasic HLH and bHLH proteins is suggested by the properties of two pairs of HLH proteins: the *da-Ac-S/emc* pair and the MyoD/Id pair. A model for their functions in forming a regulatory network is illustrated in **Figure 22.27**.

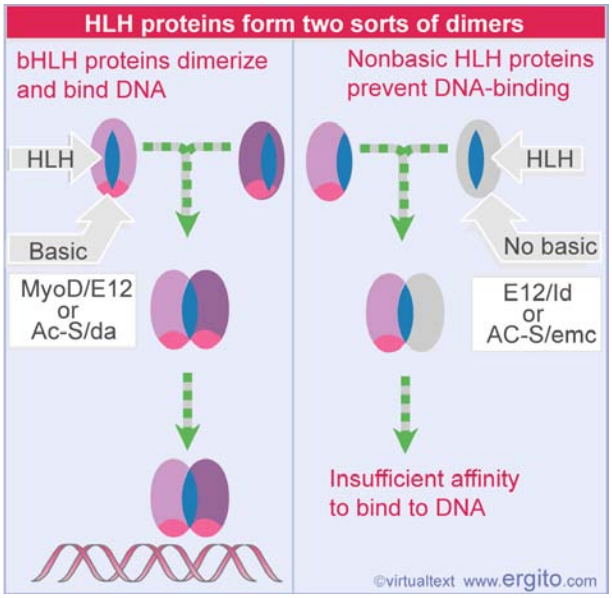

Figure 22.27 An HLH dimer in which both subunits are of the bHLH type can bind DNA, but a dimer in which one subunit lacks the basic region cannot bind DNA.

In *D. melanogaster*, the gene *emc (extramacrochaetae)* is required to establish the normal spatial pattern of adult sensory organs. It functions by *suppressing* the functions of several genes, including *da* (*daughterless*) and the *achaete-scute* complex (*Ac-S*). *Ac-S* and *da* are genes of the bHLH type. The suppressor *emc* codes for an HLH protein that lacks the basic region. We suppose that, in the absence of *emc* function, the *da* and *Ac-S* proteins form dimers that activate transcription of appropriate target genes, but the production of *emc* protein causes the formation of heterodimers that cannot bind to DNA. So production of *emc* protein in the appropriate cells is necessary to suppress the function of *Ac-S/da*.

The formation of muscle cells is triggered by a change in the transcriptional program that requires several bHLH proteins, including MyoD. MyoD is produced specifically in myogenic cells; and, indeed, overexpression of MyoD in certain other cells can induce them to commence a myogenic program. The trigger for muscle differentiation is probably a heterodimer consisting of MyoD-E12 or MyoD-E47, rather than a MyoD homodimer. Before myogenesis begins, a member of the nonbasic HLH type, the Id protein, may bind to MyoD and/or E12 and E47 to form heterodimers that cannot bind to DNA. It binds to E12/E47 better than to MyoD, and so might function by sequestering the ubiquitous bHLH partner. Overexpression of Id can prevent myogenesis. So the removal of Id could be the trigger that releases

MyoD to initiate myogenesis (668; 669; for review see 236).

A bHLH activator such as MyoD can be controlled in several ways. It is prevented from binding to DNA when it is sequestered by an HLH partner such as Id. It can activate transcription when bound to bHLH partner such as E12 or E47. It can also act as a site-specific repressor when bound to another partner; the bHLH protein MyoR forms a MyoD-MyoR dimer in proliferating myoblasts that represses transcription (at the same target loci at which MyoD-E12/E47 activate transcription).

The behavior of the HLH proteins therefore illustrates two general principles of transcriptional regulation. A small number of proteins form combinatorial associations. Particular combinations have different functions with regard to DNA binding and transcriptional regulation. Differentiation may depend either on the presence or on the removal of particular partners.

Reviews

236. Weintraub, H. (1991). *The MyoD gene family: nodal point during specification of the muscle cell lineage*. Science 251, 761-766.

References

668. Davis, R. L. et al. (1987). *Expression of a single transfected cDNA converts fibroblasts to myoblasts*. Cell 51, 987-1000.

669. Davis, R. L. et al. (1990). *The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation*. Cell 60, 733-746.

670. Benezra, R. et al. (1990). *The protein Id: a negative regulator of helix-loop-helix DNA-binding proteins*. Cell 61, 49-59.

671. Lassar, A. B. et al. (1991). *Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vitro*. Cell 66, 305-315.

675. Murre, C., McCaw, P. S., and Baltimore, D. (1989). *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins*. Cell 56, 777-783.

ACTIVATING TRANSCRIPTION

5.22.16 Leucine zippers are involved in dimer formation

Key Terms

The **leucine zipper** is a dimerization motif adjacent to a basic DNA-binding region that is found in a class of transcription factors.

A **bZIP** protein has a basic DNA-binding region adjacent to a leucine zipper dimerization motif.

Key Concepts

- The leucine zipper is an amphipathic helix that dimerizes.
- The zipper is adjacent to a basic region that binds DNA.
- Dimerization forms the bZIP motif in which the two basic regions symmetrically bind inverted repeats in DNA.

Interactions between proteins are a common theme in building a transcription complex, and a motif found in several activators (and other proteins) is involved in both homo- and heteromeric interactions. The **leucine zipper** is a stretch of amino acids rich in leucine residues that provide a dimerization motif. Dimer formation itself has emerged as a common principle in the action of proteins that recognize specific DNA sequences, and in the case of the leucine zipper, its relationship to DNA binding is especially clear, because we can see how dimerization juxtaposes the DNA-binding regions of each subunit. The reaction is depicted diagrammatically in **Figure 22.28** (673).

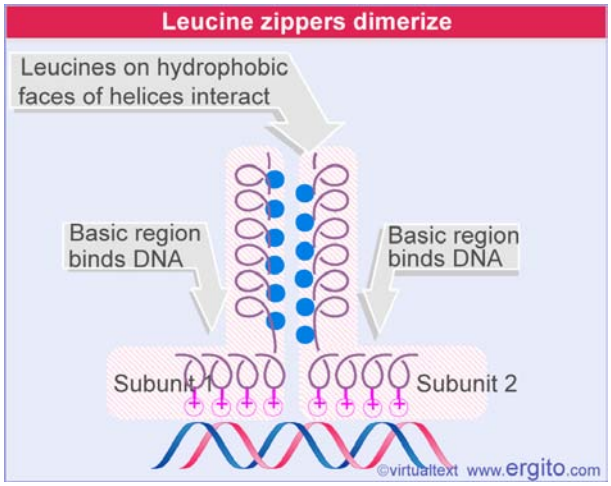

Figure 22.28 The basic regions of the bZIP motif are held together by the dimerization at the adjacent zipper region when the hydrophobic faces of two leucine zippers interact in parallel orientation.

An amphipathic α-helix has a structure in which the hydrophobic groups (including leucine) face one side, while charged groups face the other side. A leucine zipper forms an amphipathic helix in which the leucines of the zipper on one protein could protrude from the α-helix and interdigitate with the leucines of the zipper of another protein in parallel to form a coiled coil. The two right-handed helices wind around each other, with 3.5 residues per turn, so the pattern repeats integrally every 7 residues.

How is this structure related to DNA binding? The region adjacent to the leucine repeats is highly basic in each of the zipper proteins, and could comprise a DNA-binding site. The two leucine zippers in effect form a Y-shaped structure, in which the zippers comprise the stem, and the two basic regions stick out to form the arms that bind to DNA. This is known as the **bZIP** structural motif. It explains why the target sequences for such proteins are inverted repeats with no separation (676).

Zippers may be used to sponsor formation of homodimers or heterodimers. They are lengthy motifs. Leucine (or another hydrophobic amino acid) occupies every seventh residue in the potential zipper. There are 4 repeats of the zipper (Leu-X_6) in the protein C/EBP (a factor that binds as a dimer to both the CAAT box and the SV40 core enhancer), and 5 repeats in the factors Jun and Fos (which form the heterodimeric activator, AP1).

AP1 was originally identified by its binding to a DNA sequence in the SV40 enhancer (see **Figure 21.24**). The active preparation of AP1 includes several polypeptides. A major component is Jun, the product of the gene *c-jun*, which was identified by its relationship with the oncogene *v-jun* carried by an avian sarcoma virus (see *Molecular Biology 6.30.18 Oncoproteins may regulate gene expression*). The mouse genome contains a family of related genes, *c-jun* (the original isolate) and *junB* and *junD* (identified by sequence homology with *jun)*. There are considerable sequence similarities in the three Jun proteins; they have leucine zippers that can interact to form homodimers or heterodimers.

The other major component of AP1 is the product of another gene with an oncogenic counterpart. The *c-fos* gene is the cellular homologue to the oncogene *v-fos* carried by a murine sarcoma virus. Expression of *c-fos* activates genes whose promoters or enhancers possess an AP1 target site. The *c-fos* product is a nuclear phosphoprotein that is one of a group of proteins. The others are described as Fos-related antigens (FRA); they constitute a family of Fos-like proteins.

Fos also has a leucine zipper. Fos cannot form homodimers, but can form a heterodimer with Jun. A leucine zipper in each protein is required for the reaction. The ability to form dimers is a crucial part of the interaction of these factors with DNA. Fos cannot by itself bind to DNA, possibly because of its failure to form a dimer. But the Jun-Fos heterodimer can bind to DNA with same target specificity as the Jun-Jun dimer; and this heterodimer binds to the AP1 site with an affinity ~10× that of the Jun homodimer.

Reviews

676. Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989). *Scissors-grip model for DNA recognition by a family of leucine zipper proteins*. Science 246, 911-916.

References

673. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988). *The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins*. Science 240, 1759-1764.

ACTIVATING TRANSCRIPTION

5.22.17 Summary

Transcription factors include basal factors, activators, and coactivators. Basal factors interact with RNA polymerase at the startpoint. Activators bind specific short response elements (REs) located in promoters or enhancers. Activators function by making protein-protein interactions with the basal apparatus. Some activators interact directly with the basal apparatus; others require coactivators to mediate the interaction. Activators often have a modular construction, in which there are independent domains responsible for binding to DNA and for activating transcription. The main function of the DNA-binding domain may be to tether the activating domain in the vicinity of the initiation complex. Some response elements are present in many genes and are recognized by ubiquitous factors; others are present in a few genes and are recognized by tissue-specific factors.

Promoters for RNA polymerase II contain a variety of short *cis*-acting elements, each of which is recognized by a *trans*-acting factor. The *cis*-acting elements are located upstream of the TATA box and may be present in either orientation and at a variety of distances with regard to the startpoint. The upstream elements are recognized by activators that interact with the basal transcription complex to determine the efficiency with which the promoter is used. Some activators interact directly with components of the basal apparatus; others interact via intermediaries called coactivators. The targets in the basal apparatus are the TAFs of $TF_{II}D$, or $TF_{II}B$ or $TF_{II}A$. The interaction stimulates assembly of the basal apparatus.

Several groups of transcription factors have been identified by sequence homologies. The homeodomain is a 60 residue sequence found in genes that regulate development in insects and worms and in mammalian transcription factors. It is related to the prokaryotic helix-turn-helix motif and provides the motif by which the factors bind to DNA.

Another motif involved in DNA-binding is the zinc finger, which is found in proteins that bind DNA or RNA (or sometimes both). A finger has cysteine residues that bind zinc. One type of finger is found in multiple repeats in some transcription factors; another is found in single or double repeats in others.

Steroid receptors were the first members identified of a group of transcription factors in which the protein is activated by binding a small hydrophobic hormone. The activated factor becomes localized in the nucleus, and binds to its specific response element, where it activates transcription. The DNA-binding domain has zinc fingers. The receptors are homodimers or heterodimers. The homodimers all recognize palindromic response elements with the same consensus sequence; the difference between the response elements is the spacing between the inverted repeats. The heterodimers recognize direct repeats, again being distinguished by the spacing between the repeats. The DNA-binding motif of these receptors includes two zinc fingers; the first determines which consensus sequence is recognized, and the second responds to the spacing between the repeats.

The leucine zipper contains a stretch of amino acids rich in leucine that are involved

in dimerization of transcription factors. An adjacent basic region is responsible for binding to DNA.

HLH (helix-loop-helix) proteins have amphipathic helices that are responsible for dimerization, adjacent to basic regions that bind to DNA. bHLH proteins have a basic region that binds to DNA, and fall into two groups: ubiquitously expressed and tissue-specific. An active protein is usually a heterodimer between two subunits, one from each group. When a dimer has one subunit that does not have the basic region, it fails to bind DNA, so such subunits can prevent gene expression. Combinatorial associations of subunits form regulatory networks.

Many transcription factors function as dimers, and it is common for there to be multiple members of a family that form homodimers and heterodimers. This creates the potential for complex combinations to govern gene expression. In some cases, a family includes inhibitory members, whose participation in dimer formation prevents the partner from activating transcription.