PROMOTERS AND ENHANCERS

5.21.1 Introduction

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.

The **basal transcription apparatus** is the complex of transcription factors that assembles at the promoter before RNA polymerase is bound.

An **enhancer** is a *cis*-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter.

Initiation of transcription requires the enzyme RNA polymerase and transcription factors. Any protein that is needed for the initiation of transcription, but which is not itself part of RNA polymerase, is defined as a transcription factor. Many transcription factors act by recognizing *cis*-acting sites on DNA. However, binding to DNA is not the only means of action for a transcription factor. A factor may recognize another factor, or may recognize RNA polymerase, or may be incorporated into an initiation complex only in the presence of several other proteins. The ultimate test for membership of the transcription apparatus is functional: a protein must be needed for transcription to occur at a specific promoter or set of promoters.

A significant difference between the transcription of eukaryotic and prokaryotic mRNAs is that initiation at a eukaryotic promoter involves a large number of factors that bind to a variety of *cis*-acting elements. The promoter is defined as the region containing all these binding sites, that is, which can support transcription at the normal efficiency and with the proper control. So the major feature defining the promoter for a eukaryotic mRNA is the location of binding sites for transcription factors. RNA polymerase itself binds around the startpoint, but does not directly contact the extended upstream region of the promoter. By contrast, the bacterial promoters discussed in *Molecular Biology 3.9 Transcription* are largely defined in terms of the binding site for RNA polymerase in the immediate vicinity of the startpoint.

Transcription in eukaryotic cells is divided into three classes. Each class is transcribed by a different RNA polymerase:

- RNA polymerase I transcribes rRNA
- RNA polymerase II transcribes mRNA
- RNA polymerase III transcribes tRNA and other small RNAs.

Transcription factors are needed for initiation, but are not required subsequently. For

the three eukaryotic enzymes, the *factors*, rather than the RNA polymerases themselves, are principally responsible for recognizing the promoter. This is different from bacterial RNA polymerase, where it is the enzyme that recognizes the promoter sequences. For all eukaryotic RNA polymerases, the factors create a structure at the promoter to provide the target that is recognized by the enzyme. For RNA polymerases I and III, these factors are relatively simple, but for RNA polymerase II they form a sizeable group collectively known as the **basal factors**. The basal factors join with RNA polymerase II to form a complex surrounding the startpoint, and they determine the site of initiation. The basal factors together with RNA polymerase constitute the **basal transcription apparatus**.

The promoters for RNA polymerases I and II are (mostly) upstream of the startpoint, but some promoters for RNA polymerase III lie downstream of the startpoint. Each promoter contains characteristic sets of short conserved sequences that are recognized by the appropriate class of factors. RNA polymerases I and III each recognize a relatively restricted set of promoters, and rely upon a small number of accessory factors.

Promoters utilized by RNA polymerase II show more variation in sequence, and have a modular organization. Short sequence elements that are recognized by transcription factors lie upstream of the startpoint. These *cis*-acting sites usually are spread out over a region of >200 bp. Some of these elements and the factors that recognize them are common: they are found in a variety of promoters and are used constitutively. Others are specific: they identify particular classes of genes and their use is regulated. The elements occur in different combinations in individual promoters.

All RNA polymerase II promoters have sequence elements close to the startpoint that are bound by the basal apparatus and that establish the site of initiation. The sequences farther upstream determine whether the promoter is expressed in all cell types or is specifically regulated. Promoters that are constitutively expressed (their genes are sometimes called housekeeping genes) have upstream sequence elements that are recognized by ubiquitous activators. No element/factor combination is an essential component of the promoter, which suggests that initiation by RNA polymerase II may be sponsored in many different ways. Promoters that are expressed only in certain times or places have sequence elements that require activators that are available only at those times or places.

Sequence components of the promoter are defined operationally by the demand that they must be located in the general vicinity of the startpoint and are required for initiation. The **enhancer** is another type of site involved in initiation. It is identified by sequences that stimulate initiation, but that are located a considerable distance from the startpoint. Enhancer elements are often targets for tissue-specific or temporal regulation. **Figure 21.1** illustrates the general properties of promoters and enhancers.

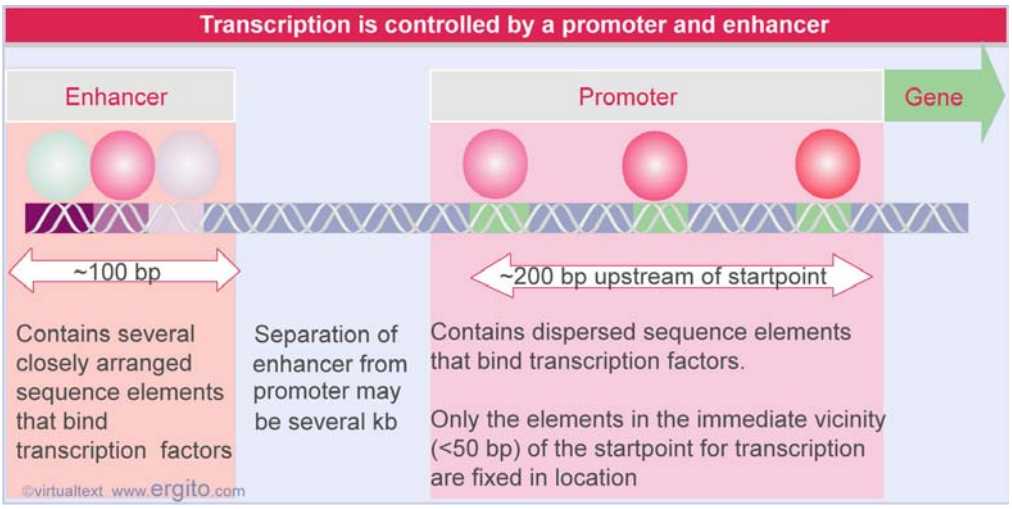

Figure 21.1 A typical gene transcribed by RNA polymerase II has a promoter that extends upstream from the site where transcription is initiated. The promoter contains several short (<10 bp) sequence elements that bind transcription factors, dispersed over >200 bp. An enhancer containing a more closely packed array of elements that also bind transcription factors may be located several kb distant. (DNA may be coiled or otherwise rearranged so that transcription factors at the promoter and at the enhancer interact to form a large protein complex.)

The components of an enhancer resemble those of the promoter; they consist of a variety of modular elements. However, the elements are organized in a closely packed array. The elements in an enhancer function like those in the promoter, but the enhancer does not need to be near the startpoint. However, proteins bound at enhancer elements interact with proteins bound at promoter elements. The distinction between promoters and enhancers is operational, rather than implying a fundamental difference in mechanism. This view is strengthened by the fact that some types of element are found in both promoters and enhancers.

Eukaryotic transcription is most often under positive regulation: a transcription factor is provided under tissue-specific control to activate a promoter or set of promoters that contain a common target sequence. Regulation by specific repression of a target promoter is less common.

A eukaryotic transcription unit generally contains a single gene, and termination occurs beyond the end of the coding region. Termination lacks the regulatory importance that applies in prokaryotic systems. RNA polymerases I and III terminate at discrete sequences in defined reactions, but the mode of termination by RNA polymerase II is not clear. However, the significant event in generating the 3′ end of an mRNA is not the termination event itself, but instead results from a cleavage reaction in the primary transcript (see *Molecular Biology 5.24 RNA splicing and processing*).

Last updated on 4-30-2001

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.21.1

PROMOTERS AND ENHANCERS

5.21.2 Eukaryotic RNA polymerases consist of many subunits

Key Terms

The **carboxy terminal domain (CTD)** of eukaryotic RNA polymerase is phosphorylated at initiation and is involved in coordinating several activities with transcription.

Amanitin (more fully α -amanitin) is a bicyclic octapeptide derived from the poisonous mushroom *Amanita phalloides*; it inhibits transcription by certain eukaryotic RNA polymerases, especially RNA polymerase II.

Key Concepts

- RNA polymerase I synthesizes rRNA in the nucleolus.
- RNA polymerase II synthesizes mRNA in the nucleoplasm.
- RNA polymerase III synthesizes small RNAs in the nucleoplasm.
- All eukaryotic RNA polymerases have ~12 subunits and are aggregates of >500 kD.
- Some subunits are common to all three RNA polymerases.
- The largest subunit in RNA polymerase II has a CTD (carboxy-terminal domain) consisting of multiple repeats of a septamer.

The three eukaryotic RNA polymerases have different locations in the nucleus, corresponding with the genes that they transcribe (for review see 71).

The most prominent activity is the enzyme RNA polymerase I, which resides in the nucleolus and is responsible for transcribing the genes coding for rRNA. It accounts for most cellular RNA synthesis (in terms of quantity).

The other major enzyme is RNA polymerase II, located in the nucleoplasm (the part of the nucleus excluding the nucleolus). It represents most of the remaining cellular activity and is responsible for synthesizing heterogeneous nuclear RNA (hnRNA), the precursor for mRNA.

RNA polymerase III is a minor enzyme activity. This nucleoplasmic enzyme synthesizes tRNAs and other small RNAs.

All eukaryotic RNA polymerases are large proteins, appearing as aggregates of >500 kD. They typically have ~12 subunits. The purified enzyme can undertake template-dependent transcription of RNA, but is not able to initiate selectively at promoters. The general constitution of a eukaryotic RNA polymerase II enzyme as

typified in *S. cerevisiae* is illustrated in **Figure 21.2**. The two largest subunits are homologous to the β and β ′ subunits of bacterial RNA polymerase. Three of the remaining subunits are common to all the RNA polymerases, that is, they are also components of RNA polymerases I and III (for review see 222).

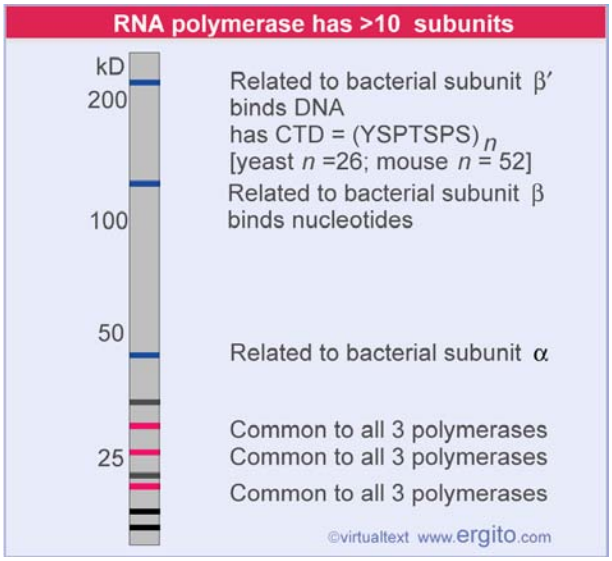

Figure 21.2 Some subunits are common to all classes of eukaryotic RNA polymerases and some are related to bacterial RNA polymerase.

The largest subunit in RNA polymerase II has a carboxy-terminal domain (**CTD**), which consists of multiple repeats of a consensus sequence of 7 amino acids. The sequence is unique to RNA polymerase II. There are ~26 repeats in yeast and ~50 in mammals. The number of repeats is important, because deletions that remove (typically) more than half of the repeats are lethal (in yeast). The CTD can be highly phosphorylated on serine or threonine residues; this is involved in the initiation reaction (see *Molecular Biology 5.21.11 Initiation is followed by promoter clearance*).

The RNA polymerases of mitochondria and chloroplasts are smaller, and resemble bacterial RNA polymerase rather than any of the nuclear enzymes. Of course, the organelle genomes are much smaller, the resident polymerase needs to transcribe relatively few genes, and the control of transcription is likely to be very much simpler (if existing at all). So these enzymes are analogous to the phage enzymes that do not need the ability to respond to a more complex environment.

A major practical distinction between the eukaryotic enzymes is drawn from their response to the bicyclic octapeptide α-**amanitin**. In basically all eukaryotic cells the activity of RNA polymerase II is rapidly inhibited by low concentrations of α-amanitin. RNA polymerase I is not inhibited. The response of RNA polymerase III to α-amanitin is less well conserved; in animal cells it is inhibited by high levels, but in yeast and insects it is not inhibited.

Last updated on 4-15-2001

Reviews

71. Doi, R. H. and Wang, L.-F. (1986). *Multiple prokaryotic RNA polymerase sigma factors*. Microbiol. Rev. 50, 227-243.

222. Young, R. A. (1991). *RNA polymerase II*. Annu. Rev. Biochem. 60, 689-715.

PROMOTERS AND ENHANCERS

5.21.3 Promoter elements are defined by mutations and footprinting

Key Terms

A **transgene** is a gene that is introduced into a cell or animal from an external source.

Key Concepts

- Promoters are defined by their ability to cause transcription of an attached sequence in an appropriate test system *in vitro* or *in vivo*.

The first step in characterizing a promoter is to define the overall length of DNA that contains all the necessary sequence elements. To do this, we need a test system in which the promoter is responsible for the production of an easily assayed product. Historically, several types of systems have been used:

- In the *oocyte system*, a DNA template is injected into the nucleus of the *X. laevis* oocyte. The RNA transcript can be recovered and analyzed. The main limitation of this system is that it is restricted to the conditions that prevail in the oocyte. It allows characterization of DNA sequences, but not of the factors that normally bind them.

- *Transfection systems* allow exogenous DNA to be introduced into a cultured cell and expressed. (The procedure is discussed in *Molecular Biology 4.18.17 Transfection introduces exogenous DNA into cells*.) The system is genuinely *in vivo* in the sense that transcription is accomplished by the same apparatus responsible for expressing the cell's own genome. However, it differs from the natural situation because the template consists of a gene that would not usually be transcribed in the host cell. The usefulness of the system may be extended by using a variety of host cells.

- *Transgenic systems* involve the addition of a gene to the germline of an animal. Expression of the **transgene** can be followed in any or all of the tissues of the animal. Some common limitations apply to transgenic systems and to transfection: the additional gene often is present in multiple copies, and is integrated at a different location from the endogenous gene. Discrepancies between the expression of a gene *in vitro* and its expression as a transgene can yield important information about the role of the genomic context of the gene.

- The *in vitro* system takes the classic approach of purifying all the components and manipulating conditions until faithful initiation is seen. "Faithful" initiation is defined as production of an RNA starting at the site corresponding to the 5′ end of mRNA (or rRNA or tRNA precursors). Ultimately this allows us to characterize the individual sequence elements in the promoter and the transcription factors that bind to them.

When a promoter is analyzed, it is important that *only* the promoter sequence changes. **Figure 21.3** shows that the same long upstream sequence is always placed next to the promoter to ensure that it is always in the same context. Because termination does not occur properly in the *in vitro* systems, the template is cut at some distance from the promoter (usually ~500 bp downstream), to ensure that all polymerases "run off" at the same point, generating an identifiable transcript.

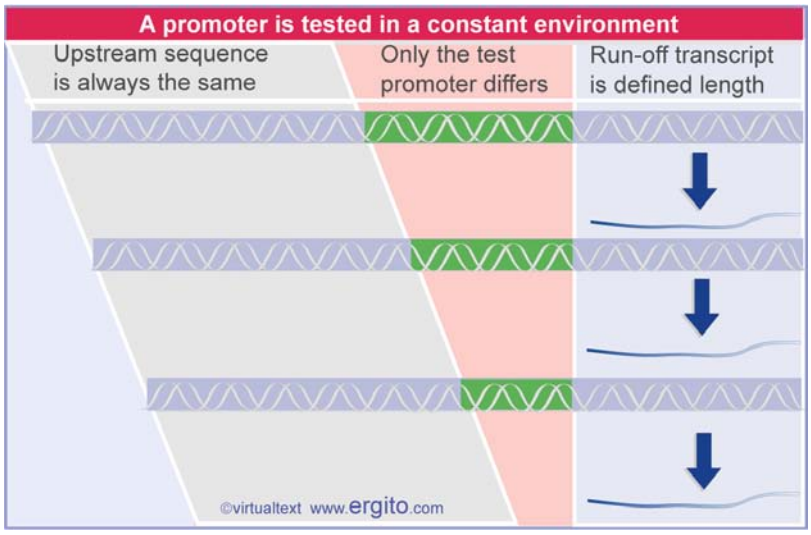

Figure 21.3 A promoter is tested by modifying the sequence that is connected to a constant upstream sequence and a constant downstream transcription unit.

We start with a particular fragment of DNA that can initiate transcription in one of these systems. Then the boundaries of the sequence constituting the promoter can be determined by reducing the length of the fragment from either end, until at some point it ceases to be active, as illustrated in **Figure 21.4**. The boundary upstream can be identified by progressively removing material from this end until promoter function is lost. To test the boundary downstream, it is necessary to reconnect the shortened promoter to the sequence to be transcribed (since otherwise there is no product to assay).

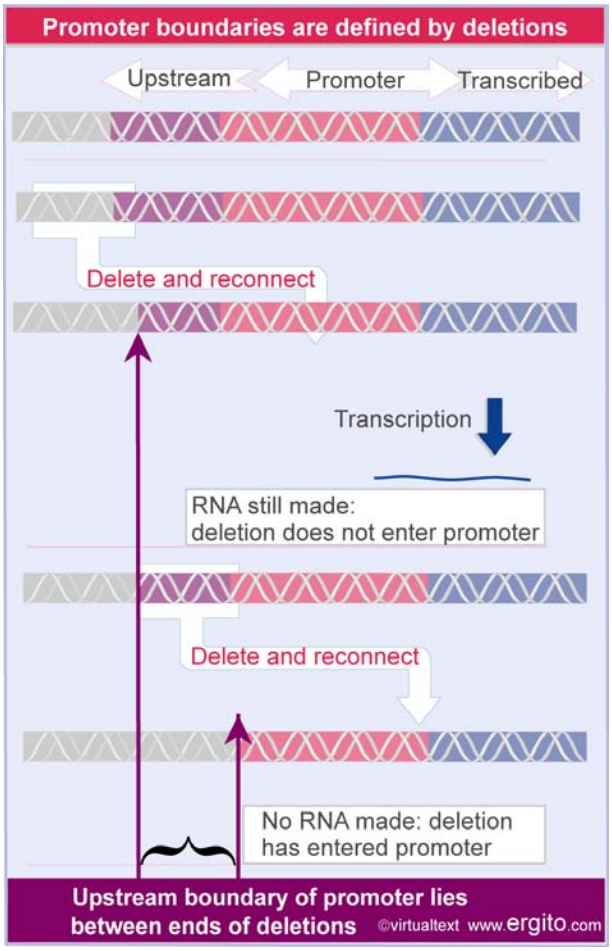

Figure 21.4 Promoter boundaries can be determined by making deletions that progressively remove more material from one side. When one deletion fails to prevent RNA synthesis but the next stops transcription, the boundary of the promoter must lie between them.

Once the boundaries of the promoter have been defined, the importance of particular bases within it can be determined by introducing point mutations or other rearrangements in the sequence. As with bacterial RNA polymerase, these can be characterized as *up* or *down* mutations. Some of these rearrangements affect only the *rate* of initiation; others influence the *site* at which initiation occurs, as seen in a change of the startpoint. To be sure that we are dealing with comparable products, in each case it is necessary to characterize the 5′ end of the RNA.

We can apply several criteria in identifying the sequence components of a promoter (or any other site in DNA):

- Mutations in the site prevent function *in vitro* or *in vivo.* (Many techniques now exist for introducing point mutations at particular base pairs, and in principle every position in a promoter can be mutated, and the mutant sequence tested *in vitro* or *in vivo.*)

- Proteins that act by binding to a site may be footprinted on it. There should be a correlation between the ability of mutations to prevent promoter function and to prevent binding of the factor.
- When a site recognized by a particular factor is present at multiple promoters, it should be possible to derive a consensus sequence that is bound by the factor. A new promoter should become responsive to this factor when an appropriate copy of the element is introduced.

PROMOTERS AND ENHANCERS

5.21.4 RNA polymerase I has a bipartite promoter

Key Terms

A **spacer** is a sequence in a gene cluster that separates the repeated copies of the transcription unit.

Key Concepts

- The RNA polymerase I promoter consists of a core promoter and an upstream control element (UPE).
- The factor UBF1 wraps DNA around a protein structure to bring the core and UPE into proximity.
- SL1 includes the factor TBP that is involved in initiation by all three RNA polymerases.
- RNA polymerase binds to the UBF1-SL1 complex at the core promoter.

RNA polymerase I transcribes only the genes for ribosomal RNA, from a single type of promoter. The transcript includes the sequences of both large and small rRNAs, which are later released by cleavages and processing. There are many copies of the transcription unit, alternating with nontranscribed **spacers**, and organized in a cluster as discussed in *Molecular Biology 1.4.8 Genes for rRNA form tandem repeats*. The organization of the promoter, and the events involved in initiation, are illustrated in **Figure 21.5**.

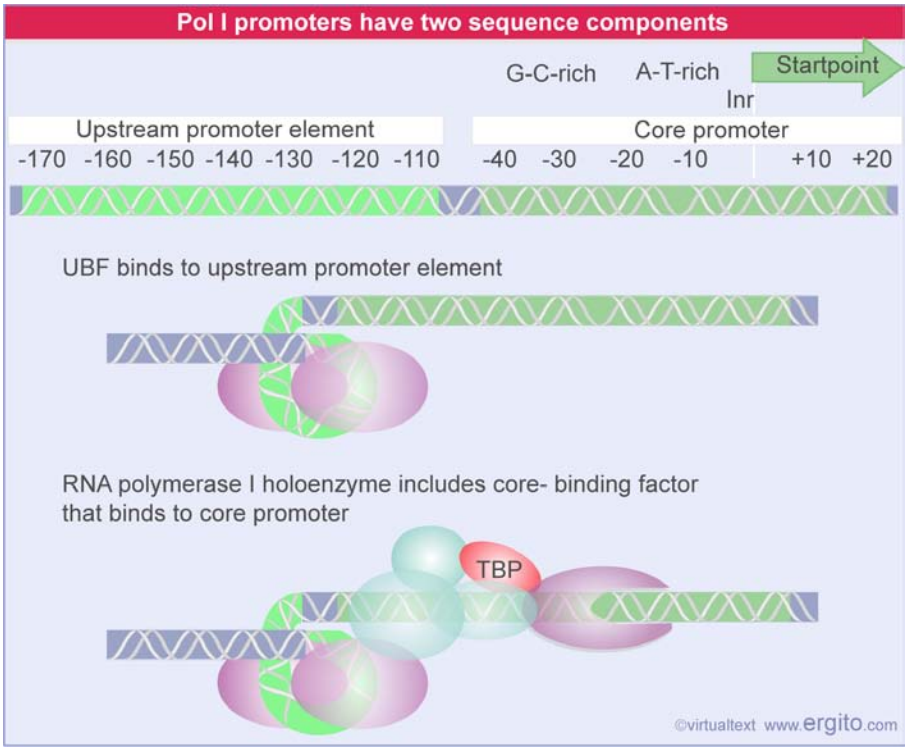

Figure 21.5 Transcription units for RNA polymerase I have a core promoter separated by ~70 bp from the upstream promoter element. UBF binding to the UPE increases the ability of core-binding factor to bind to the core promoter. Core-binding factor positions RNA polymerase I at the startpoint.

The promoter consists of two separate regions. The **core promoter** surrounds the startpoint, extending from –45 to +20, and is sufficient for transcription to initiate. It is generally G·C-rich (unusual for a promoter) except for the only conserved sequence element, a short A·T-rich sequence around the startpoint called the Inr. However, its efficiency is very much increased by the upstream promoter element (UPE), another G·C-rich sequence, related to the core promoter sequence, which extends from –180 to –107. This type of organization is common to pol I promoters in many species, although the actual sequences vary widely (for review see 1672).

RNA polymerase I requires two ancillary factors. The factor that binds to the core promoter consists of 4 proteins. (It is called SL1, TIF-IB, Rib1 in different species). One of its components, called TBP, is a factor that is required also for initiation by RNA polymerases II and III (see *Molecular Biology 5.21.8 TBP is a universal factor*). TBP does not bind directly to G·C-rich DNA, and DNA-binding is the responsibility of the other components of the core-binding factor. It is likely that TBP interacts with RNA polymerase, probably with a common subunit or a feature that has been conserved among polymerases. Core-binding factor enables RNA polymerase I to initiate from the promoter at a low basal frequency.

The core-binding factor has primary responsibility for ensuring that the RNA polymerase is properly localized at the startpoint. We see shortly that a comparable function is provided for RNA polymerases II and III by a factor that consists of TBP associated with other proteins. So a common feature in initiation by all three polymerases is a reliance on a "positioning" factor that consists of TBP associated

with proteins that are specific for each type of promoter.

For high frequency initiation, the factor UBF is required. This is a single polypeptide that binds to a G·C-rich element in the upstream promoter element. One indication of how UBF interacts with the core-binding factor is given by the importance of the spacing between the upstream promoter element and the core promoter. This can be changed by distances involving integral numbers of turns of DNA, but not by distances that introduce half turns (1676). UBF binds to the minor groove of DNA and wraps the DNA in a loop of almost 360°, thus bringing the core and the UPE into close proximity.

Figure 21.5 shows initiation as a series of sequential interactions. However, RNA polymerase I exists as a holoenzyme that contains most or all of the factors required for initiation, and which is probably recruited directly to the promoter (for review see 4177).

Last updated on 9-23-2003

Reviews

1672. Paule, M. R. and White, R. J. (2000). *Survey and summary: transcription by RNA polymerases I and III.* Nucleic Acids Res. 28, 1283-1298.

4177. Grummt, I. (2003). *Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus.* Genes Dev. 17, 1691-1702.

References

1676. Bell, S. P., Learned, R. M., Jantzen, H. M., and Tjian, R. (1988). *Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis.* Science 241, 1192-1197.

PROMOTERS AND ENHANCERS

5.21.5 RNA polymerase III uses both downstream and upstream promoters

Key Concepts

- RNA polymerase III has two types of promoters.
- Internal promoters have short consensus sequences located within the transcription unit and cause initiation to occur a fixed distance upstream.
- Upstream promoters contain three short consensus sequences upstream of the startpoint that are bound by transcription factors.

Recognition of promoters by RNA polymerase III strikingly illustrates the relative roles of transcription factors and the polymerase enzyme. The promoters fall into two general classes that are recognized in different ways by different groups of factors. The promoters for 5S and tRNA genes are *internal;* they lie downstream of the startpoint. The promoters for snRNA (small nuclear RNA) genes lie upstream of the startpoint in the more conventional manner of other promoters. In both cases, the individual elements that are necessary for promoter function consist exclusively of sequences recognized by transcription factors, which in turn direct the binding of RNA polymerase.

Before the promoter of 5S RNA genes was identified in *X. laevis*, all attempts to identify promoter sequences assumed that they would lie upstream of the startpoint. But deletion analysis showed that the 5S RNA product continues to be synthesized when the entire sequence upstream of the gene is removed!

When the deletions continue into the gene, a product very similar in size to the usual 5S RNA continues to be synthesized so long as the deletion ends before base +55. **Figure 21.6** shows that the first part of the RNA product corresponds to plasmid DNA; the second part represents the segment remaining of the usual 5S RNA sequence. But when the deletion extends past +55, transcription does not occur. So the promoter lies *downstream of position +55*, but causes RNA polymerase III to initiate transcription a more or less fixed distance upstream (641; 642).

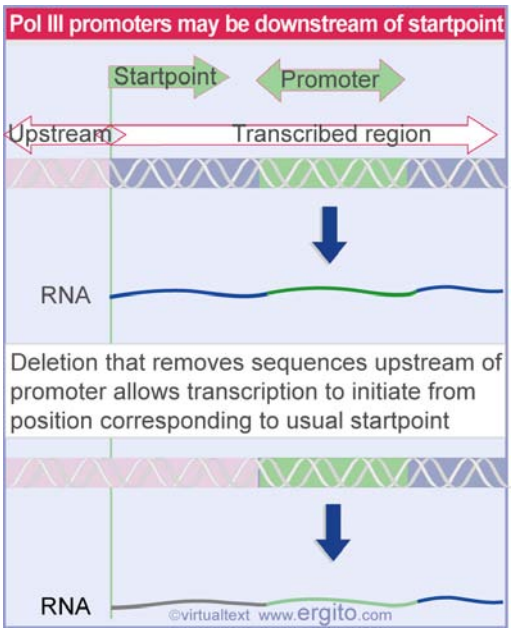

Figure 21.6 Deletion analysis shows that the promoter for 5S RNA genes is internal; initiation occurs a fixed distance (~55 bp) upstream of the promoter.

When deletions extend into the gene from its distal end, transcription is unaffected so long as the first 80 bp remain intact. Once the deletion cuts into this region, transcription ceases. This places the downstream boundary position of the promoter at about position +80.

So the promoter for 5S RNA transcription lies between positions +55 and +80 within the gene. A fragment containing this region can sponsor initiation of any DNA in which it is placed, from a startpoint ~55 bp farther upstream. (The wild-type startpoint is unique; in deletions that lack it, transcription initiates at the purine base nearest to the position 55 bp upstream of the promoter.)

The structures of three types of promoters for RNA polymerase III are summarized in **Figure 21.7**. There are two types of internal promoter. Each contains a bipartite structure, in which two short sequence elements are separated by a variable sequence. Type 1 consists of a boxA sequence separated from a boxC sequence (1673), and type 2 consists of a boxA sequence separated from a boxB sequence (1674). The distance between boxA and boxB in a type 2 promoter can vary quite extensively, but the boxes usually cannot be brought too close together without abolishing function. Type 3 promoters have three sequence elements all located upstream of the startpoint (1675).

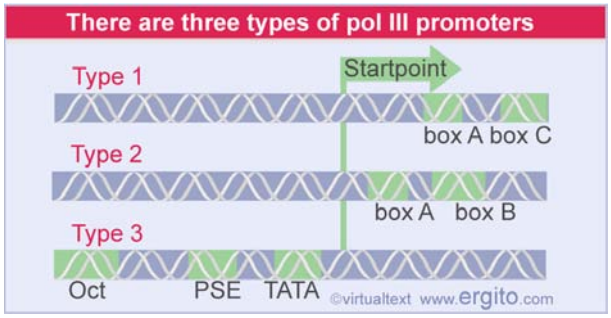

Figure 21.7 Promoters for RNA polymerase III may consist of bipartite sequences downstream of the startpoint, with boxA separated from either boxC or boxB. Or they may consist of separated sequences upstream of the startpoint (Oct, PSE, TATA).

References

641. Sakonju, S., Bogenhagen, D. F., and Brown, D. D. (1980). *A control region in the center of the 5S RNA gene directs specific initiation of transcription: I the 5′ border of the region*. Cell 19, 13-25.

642. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980). *A control region in the center of the 5S RNA gene directs specific initiation of transcription: II the 3′ border of the region*. Cell 19, 27-35.

1673. Pieler, T., Hamm, J., and Roeder, R. G. (1987). *The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing.* Cell 48, 91-100.

1674. Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981). *Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements.* Nature 294, 626-631.

1675. Kunkel, G. R. and Pederson, T. (1988). *Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used.* Genes Dev. 2, 196-204.

PROMOTERS AND ENHANCERS

5.21.6 $TF_{III}B$ is the commitment factor for pol III promoters

Key Terms

An **assembly factor** is a protein that is required for formation of a macromolecular structure but is not itself part of that structure.

Preinitiation complex in eukaryotic transcription describes the assembly of transcription factors at the promoter before RNA polymerase binds.

Key Concepts

- $TF_{III}A$ and $TF_{III}C$ bind to the consensus sequences and enable $TF_{III}B$ to bind at the startpoint.
- $TF_{III}B$ has TBP as one subunit and enables RNA polymerase to bind.

The detailed interactions are different at the two types of internal promoter, but the principle is the same. $TF_{III}C$ binds downstream of the startpoint, either independently (type 2 promoters) or in conjunction with $TF_{III}A$ (type 1 promoters). The presence of $TF_{III}C$ enables the positioning factor $TF_{III}B$ to bind at the startpoint. Then RNA polymerase is recruited.

Figure 21.8 summarizes the stages of reaction at type 2 internal promoters. $TF_{III}C$ binds to both boxA and boxB. This enables $TF_{III}B$ to bind at the startpoint. Then RNA polymerase III can bind.

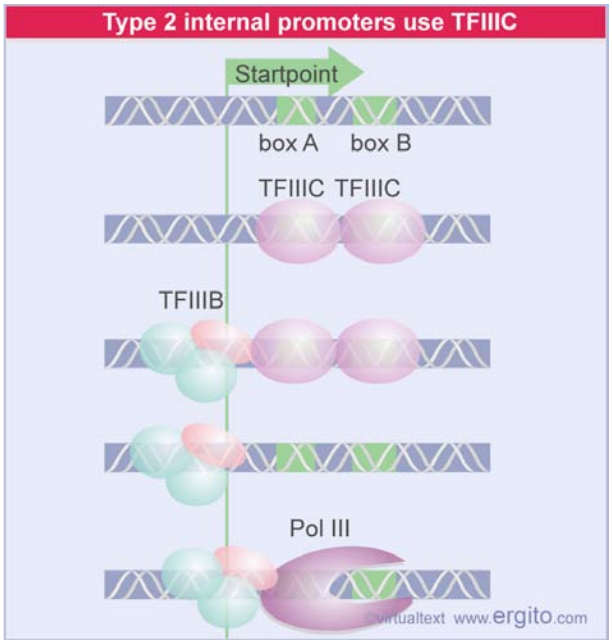

Figure 21.8 Internal type 2 pol III promoters use binding of TFIIIC to boxA and boxB sequences to recruit the positioning factor TFIIIB, which recruits RNA polymerase III.

The difference at type 1 internal promoters is that $TF_{III}A$ must bind at boxA to enable $TF_{III}C$ to bind at boxC. **Figure 21.9** shows that, once $TF_{III}C$ has bound, events follow the same course as at type 2 promoters, with $TF_{III}B$ binding at the startpoint, and RNA polymerase III joining the complex. Type 1 promoters are found only in the genes for 5S rRNA.

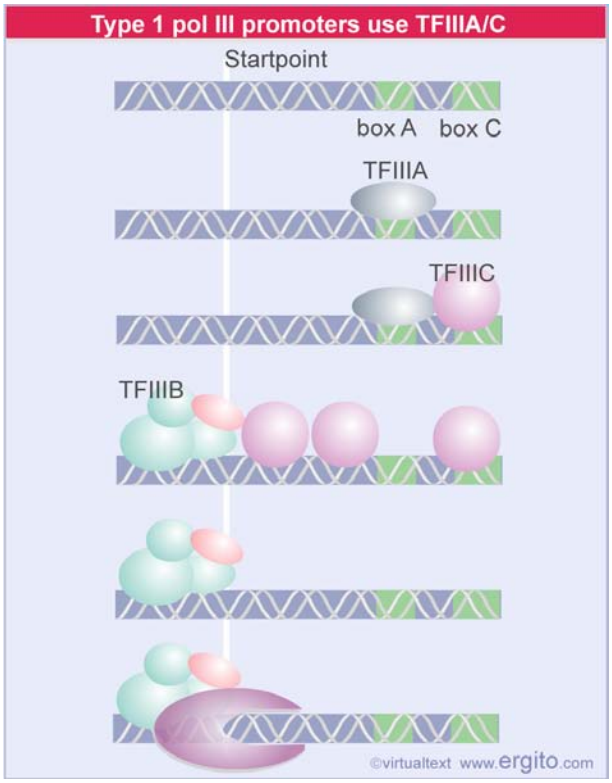

Figure 21.9 Internal type 1 pol III promoters use the assembly factors TFIIIA and TFIIIC, at boxA and boxC, to recruit the positioning factor TFIIIB, which recruits RNA polymerase III.

$TF_{III}A$ and $TF_{III}C$ are **assembly factors**, whose sole role is to assist the binding of $TF_{III}B$ at the right location. Once $TF_{III}B$ has bound, $TF_{III}A$ and $TF_{III}C$ can be removed from the promoter (by high salt concentration *in vitro*) without affecting the initiation reaction. *$TF_{III}B$ remains bound in the vicinity of the startpoint and its presence is sufficient to allow RNA polymerase III to bind at the startpoint.* So $TF_{III}B$ is the only true initiation factor required by RNA polymerase III (643). This sequence of events explains how the promoter boxes downstream can cause RNA polymerase to bind at the startpoint, farther upstream. Although the ability to transcribe these genes is conferred by the internal promoter, changes in the region immediately upstream of the startpoint can alter the efficiency of transcription.

$TF_{III}C$ is a large protein complex (>500 kD), comparable in size to RNA polymerase itself, and containing 6 subunits. $TF_{III}A$ is a member of an interesting class of proteins containing a nucleic acid-binding motif called a zinc finger (see *Molecular Biology 5.22.9 A zinc finger motif is a DNA-binding domain*). The positioning factor, $TF_{III}B$, consists of three subunits. It includes the same protein, TBP, that is present in the core-binding factor for pol I promoters, and also in the corresponding transcription factor ($TF_{II}D$) for RNA polymerase II (1677). It also contains Brf, which is related to the factor $TF_{II}B$ that is used by RNA polymerase II. The third subunit is called B′′; it is dispensable if the DNA duplex is partially melted, which suggests that its function is to initiate the transcription bubble (945). The role of B′′ may be comparable to the role played by sigma factor in bacterial RNA polymerase (see *Molecular Biology 3.9.16 Substitution of sigma factors may control initiation*).

The upstream region has a conventional role in the third class of polymerase III promoters. In the example shown in **Figure 21.7**, there are three upstream elements. These elements are also found in promoters for snRNA genes that are transcribed by RNA polymerase II. (Genes for some snRNAs are transcribed by RNA polymerase II, while others are transcribed by RNA polymerase III.) The upstream elements function in a similar manner in promoters for both polymerases II and III.

Initiation at an upstream promoter for RNA polymerase III can occur on a short region that immediately precedes the startpoint and contains only the TATA element. However, efficiency of transcription is much increased by the presence of the PSE and OCT elements. The factors that bind at these elements interact cooperatively. (The PSE element may be essential at promoters used by RNA polymerase II, whereas it is stimulatory in promoters used by RNA polymerase III; its name stands for proximal sequence element.)

The TATA element confers specificity for the type of polymerase (II or III) that is recognized by an snRNA promoter. It is bound by a factor that includes the TBP, which actually recognizes the sequence in DNA. The TBP is associated with other proteins, which are specific for the type of promoter. The function of TBP and its associated proteins is to position the RNA polymerase correctly at the startpoint. We discuss this in more detail for RNA polymerase II (see *Molecular Biology 5.21.8 TBP is a universal factor*).

The factors work in the same way for both types of promoters for RNA polymerase III. *The factors bind at the promoter before RNA polymerase itself can bind.* They form a **preinitiation complex** that directs binding of the RNA polymerase. RNA polymerase III does not itself recognizes the promoter sequence, but binds adjacent to factors that are themselves bound just upstream of the startpoint. For the type 1 and type 2 internal promoters, the assembly factors ensure that $TF_{III}B$ (which includes TBP) is bound just upstream of the startpoint, to provide the positioning information. For the upstream promoters, $TF_{III}B$ binds directly to the region including the TATA box. So irrespective of the location of the promoter sequences, factor(s) are bound close to the startpoint in order to direct binding of RNA polymerase III (for review see 220; 3231).

Last updated on 1-6-2003

Reviews

220. Geiduschek, E. P. and Tocchini-Valentini, G. P. (1988). *Transcription by RNA polymerase III.* Annu. Rev. Biochem. 57, 873-914.

3231. Schramm, L. and Hernandez, N. (2002). *Recruitment of RNA polymerase III to its target promoters.* Genes Dev. 16, 2593-2620.

References

643. Kassavatis, G. A., Braun, B. R., Nguyen, L. H., and Geiduschek, E. P. (1990). *S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors*. Cell 60, 235-245.

945. Kassavetis, G. A., Letts, G. A., and Geiduschek, E. P. (1999). *A minimal RNA polymerase III transcription system.* EMBO J. 18, 5042-5051.

1677. Kassavetis, G. A., Joazeiro, C. A., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S., and Blanco, J. A. (1992). *The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB.* Cell 71, 1055-1064.

PROMOTERS AND ENHANCERS

5.21.7 The startpoint for RNA polymerase II

Key Terms

TAFs ($TF_{II}X$) are the subunits of $TF_{II}D$ that assist TBP in binding to DNA. They also provide points of contact for other components of the transcription apparatus.

The **core promoter** of RNA polymerase I is the region immediately surrounding the startpoint. It is necessary and sufficient to initiate transcription, but only at a low level.

A **core promoter** is the shortest sequence at which an RNA polymerase can initiate transcription (typically at much lower level than that displayed by a promoter containing additional elements). For RNA polymerase II it is the minimal sequence at which the basal transcription apparatus can assemble, and includes two sequence elements, the InR and TATA box. It is typically ~40 bp long.

The **Inr** is the sequence of a pol II promoter between –3 and +5 and has the general sequence Py_2CAPy_5. It is the simplest possible pol II promoter.

TATA box is a conserved A·T-rich septamer found about 25 bp before the startpoint of each eukaryotic RNA polymerase II transcription unit; may be involved in positioning the enzyme for correct initiation.

A **TATA-less promoter** does not have a TATA box in the sequence upstream of its startpoint.

Key Concepts

- RNA polymerase II requires general transcription factors (called **$TF_{II}X$**) to initiate transcription.
- RNA polymerase II promoters have a short conserved sequence Py_2CAPy_5 (the initiator InR) at the startpoint.
- The TATA box is a common component of RNA polymerase II promoters and consists of an A·T-rich octamer located ~25 bp upstream of the startpoint
- The DPE is a common component of RNA polymerase II promoters that do not contain a TATA box.
- A core promoter for RNA polymerase II includes the InR and either a TATA box or a DPE.

The basic organization of the apparatus for transcribing protein-coding genes was revealed by the discovery that purified RNA polymerase II can catalyze synthesis of mRNA, but cannot initiate transcription unless an additional extract is added (2406). The purification of this extract led to the definition of the general transcription factors – a group of proteins that are needed for initiation by RNA polymerase II at all promoters (for review see 2407). RNA polymerase II in conjunction with these factors constitutes the basal transcription apparatus that is needed to transcribe any promoter. The general factors are described as **$TF_{II}X$**, where "X" is a letter that

identifies the individual factor. The subunits of RNA polymerase II and the general transcription factors are conserved among eukaryotes.

Our starting point for considering promoter organization is to define the **core promoter** as the shortest sequence at which RNA polymerase II can initiate transcription. A core promoter can in principle be expressed in any cell. It comprises the minimum sequence that enables the general transcription factors to assemble at the startpoint. They are involved in the mechanics of binding to DNA and enable RNA polymerase II to initiate transcription. A core promoter functions at only a low efficiency. Other proteins, called activators, are required for a proper level of function (see *Molecular Biology 5.21.13 Short sequence elements bind activators*). The activators are not described systematically, but have casual names reflecting their histories of identification.

We may expect any sequence components involved in the binding of RNA polymerase and general transcription factors to be conserved at most or all promoters. As with bacterial promoters, when promoters for RNA polymerase II are compared, homologies in the regions near the startpoint are restricted to rather short sequences. These elements correspond with the sequences implicated in promoter function by mutation. **Figure 21.10** shows the construction of a typical pol II core promoter (for review see 3225, 4527).

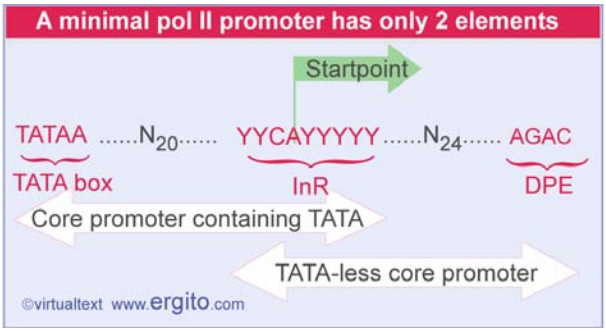

Figure 21.10 The minimal pol II promoter has a TATA box ~25 bp upstream of the InR. The TATA box has the consensus sequence of TATAA. The Inr has pyrimidines (Y) surrounding the CA at the startpoint. The sequence shows the coding strand.

At the startpoint, there is no extensive homology of sequence, but there is a tendency for the first base of mRNA to be A, flanked on either side by pyrimidines. (This description is also valid for the CAT start sequence of bacterial promoters.) This region is called the *initiator* (**Inr**), and may be described in the general form Py_2CAPy_5 (3228; 3229). The Inr is contained between positions –3 and +5.

Many promoters have a sequence called the **TATA box**, usually located ~25 bp upstream of the startpoint. It constitutes the only upstream promoter element that has a relatively fixed location with respect to the startpoint. The core sequence is TATAA, usually followed by three more A·T base pairs (see 3227). The TATA box tends to be surrounded by G·C-rich sequences, which could be a factor in its function. It is almost identical with the –10 sequence found in bacterial promoters; in fact, it could pass for one except for the difference in its location at –25 instead of –10.

Single base substitutions in the TATA box act as strong down mutations. Some mutations reverse the orientation of an A·T pair, so base composition alone is not sufficient for its function. So the TATA box comprises an element whose behavior is analogous to our concept of the bacterial promoter: a short, well-defined sequence just upstream of the startpoint, which is necessary for transcription.

Promoters that do not contain a TATA element are called **TATA-less promoters**. Surveys of promoter sequences suggest that 50% or more of promoters may be TATA-less. When a promoter does not contain a TATA box, it usually contains another element, the DPE (downstream promoter element) which is located at +28 - +32 (3230).

A core promoter can consist either of a TATA box plus InR or of an InR plus DPE.

Last updated on January 12, 2004

Reviews

2407. Woychik, N. A. and Hampsey, M. (2002). *The RNA polymerase II machinery: structure illuminates function.* Cell 108, 453-463.

3225. Butler, J. E. and Kadonaga, J. T. (2002). *The RNA polymerase II core promoter: a key component in the regulation of gene expression.* Genes Dev. 16, 2583-2592.

3229. Smale, S. T., Jain, A., Kaufmann, J., Emami, K. H., Lo, K., and Garraway, I. P. (1998). *The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes.* Cold Spring Harb Symp Quant Biol 63, 21-31.

4527. Smale, S. T. and Kadonaga, J. T. (2003). *The RNA polymerase II core promoter.* Annu. Rev. Biochem. 72, 449-479.

References

2406. Weil, P. A., Luse, D. S., Segall, J., and Roeder, R. G. (1979). *Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA.* Cell 18, 469-484.

3227. Singer, V. L., Wobbe, C. R., and Struhl, K. (1990). *A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation.* Genes Dev. 4, 636-645.

3228. Smale, S. T. and Baltimore, D. (1989). *The "initiator" as a transcription control element.* Cell 57, 103-113.

3230. Burke, T. W. and Kadonaga, J. T. (1996). *Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters.* Genes Dev. 10, 711-724.

PROMOTERS AND ENHANCERS

5.21.8 TBP is a universal factor

Key Terms

$TF_{II}D$ is the transcription factor that binds to the TATA sequence upstream of the startpoint of promoters for RNA polymerase II. It consists of TBP (TATA binding protein) and the TAF subunits that bind to TBP.

The **TATA-binding protein (TBP)** is the subunit of transcription factor $TF_{II}D$ that binds to DNA.

TAFs ($TF_{II}X$) are the subunits of $TF_{II}D$ that assist TBP in binding to DNA. They also provide points of contact for other components of the transcription apparatus.

Key Concepts

- TBP is a component of the positioning factor that is required for each type of RNA polymerase to bind its promoter.
- The factor for RNA polymerase II is $TF_{II}D$, which consists of TBP and 11 TAFs, with a total mass ~800 kD.

The first step in complex formation at a promoter containing a TATA box is binding of the factor **$TF_{II}D$** to a region that extends upstream from the TATA sequence. $TF_{II}D$ contains two types of component. Recognition of the TATA box is conferred by the **TATA-binding protein** (**TBP**), a small protein of ~30 kD. The other subunits are called **TAFs** (for TBP-associated factors). Some TAFs are stoichiometric with TBP; others are present in lesser amounts. $TF_{II}D$s containing different TAFs could recognize different promoters. Some (substoichiometric) TAFs are tissue-specific. The total mass of $TF_{II}D$ typically is ~800 kD, containing TBP and 11 TAFs, varying in mass from 30-250 kD. The TAFs in $TF_{II}D$ are named in the form $TAF_{II}00$, where "00" gives the molecular mass of the subunit.

Positioning factors that consist of TBP associated with a set of TAFs are responsible for identifying all classes of promoters. $TF_{III}B$ (for pol III promoters) and SL1 (for pol I promoters) may both be viewed as consisting of TBP associated with a particular group of proteins that substitute for the TAFs that are found in $TF_{II}D$ (for review see 1709). TBP is the key component, and is incorporated at each type of promoter by a different mechanism (for review see 2394). In the case of promoters for RNA polymerase II, the key feature in positioning is the fixed distance of the TATA box from the startpoint.

Figure 21.11 shows that the positioning factor recognizes the promoter in a different way in each case. At promoters for RNA polymerase III, $TF_{III}B$ binds adjacent to $TF_{III}C$. At promoters for RNA polymerase I, SL1 binds in conjunction with UBF. $TF_{II}D$ is solely responsible for recognizing promoters for RNA polymerase II. At a promoter that has a TATA element, TBP binds specifically to DNA, but at other promoters it may be incorporated by association with other proteins that bind to DNA. Whatever its means of entry into the initiation complex, it has the common

purpose of interaction with the RNA polymerase.

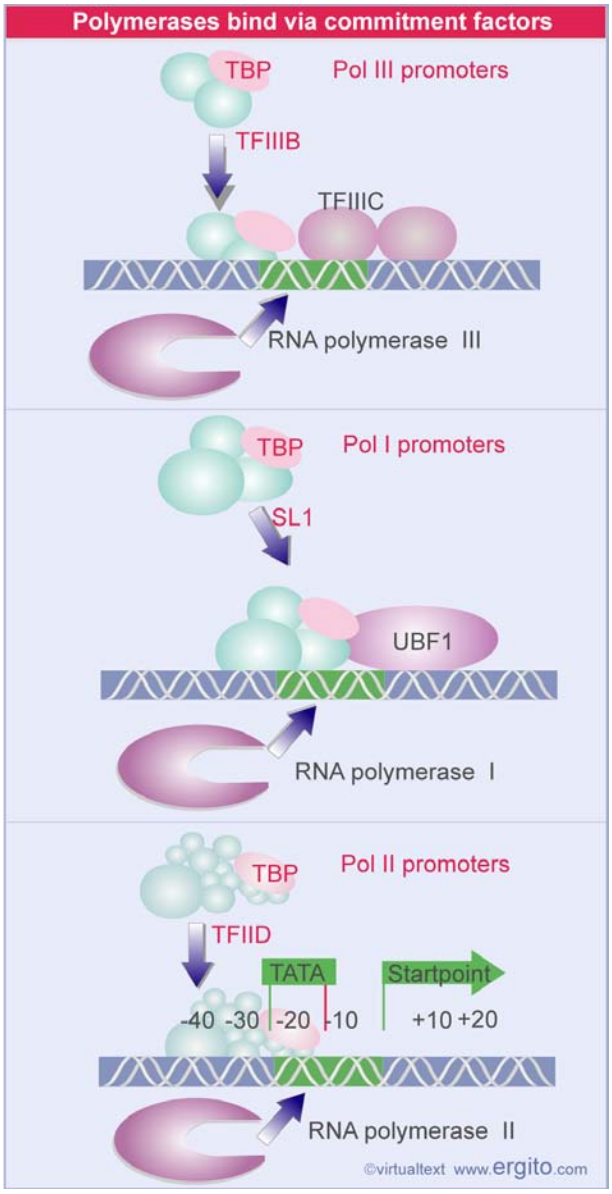

Figure 21.11 RNA polymerases are positioned at all promoters by a factor that contains TBP.

$TF_{II}D$ is ubiquitous, but not unique. All multicellular eukaryotes also express an alternative complex, which has TLF (TBP like factor) instead of TBP (1708). A TLF is typically ~60% similar to TBP. It probably initiates complex formation by the usual set of TF_{II} factors. However, TLF does not bind to the TATA box, and we do not yet know how it works. *Drosophila* also has a third factor, TRF1, which behaves in the same way as TBP and binds its own set of TAFs, to form a complex that functions as an alternative to $TF_{II}D$ at a specific set of promoters (1707).

Last updated on 4-30-2001

Reviews

1708. Berk, A. J. (2000). *TBP-like factors come into focus.* Cell 103, 5-8.

1709. Lee, T. I. and Young, R. A. (1998). *Regulation of gene expression by TBP-associated proteins.* Genes Dev. 12, 1398-1408.

2394. Hernandez, N. (1993). *TBP, a universal eukaryotic transcription factor?* Genes Dev. 7, 1291-1308.

References

1707. Crowley, T. E., Hoey, T., Liu, J. K., Jan, Y. N., Jan, L. Y., and Tjian, R. (1993). *A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila.* Nature 361, 557-561.

PROMOTERS AND ENHANCERS

5.21.9 TBP binds DNA in an unusual way

Key Concepts

- TBP binds to the TATA box in the minor groove of DNA.
- It forms a saddle around the DNA and bends it by ~80°.
- Some of the TAFs resemble histones and may form a structure resembling a histone octamer.

TBP has the unusual property of binding to DNA in the minor groove. (Virtually all known DNA-binding proteins bind in the major groove.) The crystal structure of TBP suggests a detailed model for its binding to DNA. **Figure 21.12** shows that it surrounds one face of DNA, forming a "saddle" around the double helix. In effect, the inner surface of TBP binds to DNA, and the larger outer surface is available to extend contacts to other proteins. The DNA-binding site consists of a C-terminal domain that is conserved between species, while the variable N-terminal tail is exposed to interact with other proteins (647; 648; 649). It is a measure of the conservation of mechanism in transcriptional initiation that the DNA-binding sequence of TBP is 80% conserved between yeast and Man.

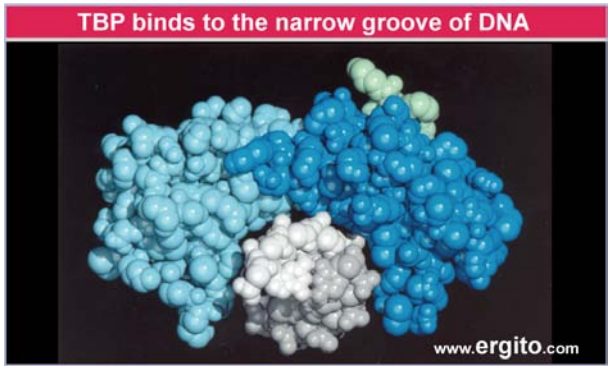

Figure 21.12 A view in cross-section shows that TBP surrounds DNA from the side of the narrow groove. TBP consists of two related (40% identical) conserved domains, which are shown in light and dark blue. The N-terminal region varies extensively and is shown in green. The two strands of the DNA double helix are in light and dark grey. Photograph kindly provided by Stephen Burley.

Binding of TBP may be inconsistent with the presence of nucleosomes. Because nucleosomes form preferentially by placing A·T-rich sequences with the minor grooves facing inward, they could prevent binding of TBP. This may explain why the presence of nucleosomes prevents initiation of transcription.

TBP binds to the minor groove and bends the DNA by ~80°, as illustrated in **Figure**

21.13. The TATA box bends towards the major groove, widening the minor groove. The distortion is restricted to the 8 bp of the TATA box; at each end of the sequence, the minor groove has its usual width of ~5 Å, but at the center of the sequence the minor groove is >9 Å. This is a deformation of the structure, but does not actually separate the strands of DNA, because base pairing is maintained. The extent of the bend can vary with the exact sequence of the TATA box, and is correlated with the efficiency of the promoter (4532).

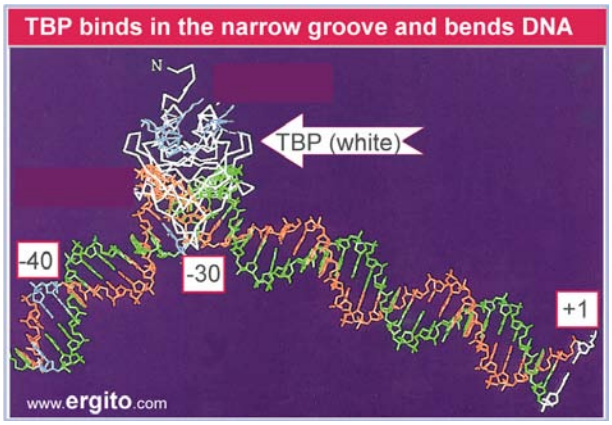

Figure 21.13 The cocrystal structure of TBP with DNA from -40 to the startpoint shows a bend at the TATA box that widens the narrow groove where TBP binds. Photograph provided by Stephen Burley.

This structure has several functional implications. By changing the spatial organization of DNA on either side of the TATA box, it allows the transcription factors and RNA polymerase to form a closer association than would be possible on linear DNA. The bending at the TATA box corresponds to unwinding of about 1/3 of a turn of DNA, and is compensated by a positive writhe.

The presence of TBP in the minor groove, combined with other proteins binding in the major groove, creates a high density of protein-DNA contacts in this region. Binding of purified TBP to DNA *in vitro* protects ~1 turn of the double helix at the TATA box, typically extending from –37 to –25; but binding of the $TF_{II}D$ complex in the initiation reaction regularly protects the region from –45 to –10, and also extends farther upstream beyond the startpoint. TBP is the only general transcription factor that makes sequence-specific contacts with DNA.

Within $TF_{II}D$ as a free protein complex, the factor $TAF_{II}230$ binds to TBP, where it occupies the concave DNA-binding surface. In fact, the structure of the binding site, which lies in the N-terminal domain of $TAF_{II}230$, mimics the surface of the minor groove in DNA. This molecular mimicry allows $TAF_{II}230$ to control the ability of TBP to bind to DNA; the N-terminal domain of $TAF_{II}230$ must be displaced from the DNA-binding surface of TBP in order for $TF_{II}D$ to bind to DNA (654).

Some TAFs resemble histones; in particular $TAF_{II}42$ and $TAF_{II}62$ appear to be (distant) homologues of histones H3 and H4, and they form a heterodimer using the same motif (the histone fold) that histones use for the interaction. (Histones H3 and H4 form the kernel of the histone octamer – the basic complex that binds DNA in

eukaryotic chromatin; see *Molecular Biology 5.20.8 Organization of the histone octamer*.) Together with other TAFs, $TAF_{II}42$ and $TAF_{II}62$ may form the basis for a structure resembling a histone octamer; such a structure may be responsible for the nonsequence-specific interactions of $TF_{II}D$ with DNA. Histone folds are also used in pairwise interactions between other $TAF_{II}s$.

Some of the $TAF_{II}s$ may be found in other complexes as well as in $TF_{II}D$. In particular, the histone-like $TAF_{II}s$ are found also in protein complexes that modify the structure of chromatin prior to transcription (see *Molecular Biology 5.23.7 Acetylases are associated with activators*) (651; 653; 657; 695; for review see 225; 1709; 2395).

Last updated on January 19, 2004

Reviews

225. Burley, S. K. and Roeder, R. G. (1996). *Biochemistry and structural biology of TFIID.* Annu. Rev. Biochem. 65, 769-799.

1709. Lee, T. I. and Young, R. A. (1998). *Regulation of gene expression by TBP-associated proteins.* Genes Dev. 12, 1398-1408.

2395. Orphanides, G., Lagrange, T., and Reinberg, D. (1996). *The general transcription factors of RNA polymerase II.* Genes Dev. 10, 2657-2683.

References

647. Nikolov, D. B. et al. (1992). *Crystal structure of TFIID TATA-box binding protein*. Nature 360, 40-46.

648. Kim, Y. et al. (1993). *Crystal structure of a yeast TBP/TATA box complex*. Nature 365, 512-520.

649. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993). *Cocrystal structure of TBP recognizing the minor groove of a TATA element*. Nature 365, 520-527.

651. Martinez, E. et al. (1994). *TATA-binding protein-associated factors in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter*. EMBO J. 13, 3115-3126.

653. Verrijzer, C. P. et al. (1995). *Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II*. Cell 81, 1115-1125.

654. Liu, D. et al. (1998). *Solution structure of a TBP-TAFII230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP*. Cell 94, 573-583.

657. Horikoshi, M. et al. (1988). *Transcription factor ATD interacts with a TATA factor to facilitate establishment of a preinitiation complex*. Cell 54, 1033-1042.

695. Ogryzko, V. V. et al. (1998). *Histone-like TAFs within the PCAF histone acetylase complex*. Cell 94, 35-44.

4532. Wu, J., Parkhurst, K. M., Powell, R. M., Brenowitz, M., and Parkhurst, L. J. (2001). *DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence-dependent.* J. Biol. Chem. 276, 14614-14622.

PROMOTERS AND ENHANCERS

5.21.10 The basal apparatus assembles at the promoter

Key Concepts

- Binding of $TF_{II}D$ to the TATA box is the first step in initiation.
- Other transcription factors bind to the complex in a defined order, extending the length of the protected region on DNA.
- When RNA polymerase II binds to the complex, it initiates transcription.

Initiation requires the transcription factors to act in a defined order to build a complex that is joined by RNA polymerase. The series of events was initially defined by following the increasing size of the protein complex associated with DNA. Now we can define the events in more detail in terms of the interactions revealed by the crystal structures of the various factors, and of RNA polymerase bound to DNA.

Footprinting of the DNA regions protected by each complex suggests the model summarized in **Figure 21.15**. As each TF_{II} factor joins the complex, an increasing length of DNA is covered. RNA polymerase is incorporated at a late stage (644; for review see 223; 226).

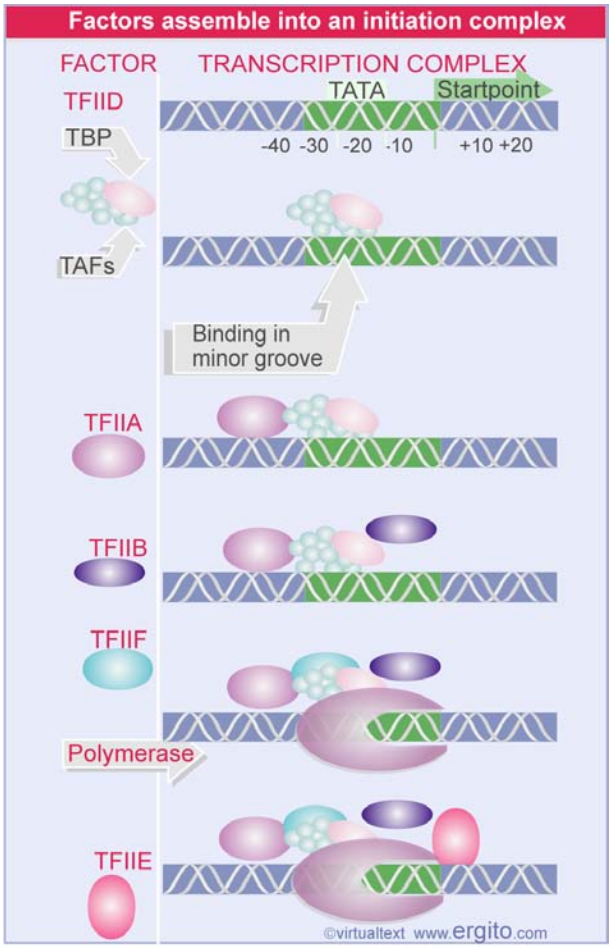

Figure 21.15 An initiation complex assembles at promoters for RNA polymerase II by an ordered sequence of association with transcription factors.

Commitment to a promoter is initiated when $TF_{II}D$ binds the TATA box. ($TF_{II}D$ also recognizes the InR sequence at the startpoint.) When $TF_{II}A$ joins the complex, $TF_{II}D$ becomes able to protect a region extending farther upstream. $TF_{II}A$ may activate TBP by relieving the repression that is caused by the $TAF_{II}230$.

Addition of $TF_{II}B$ gives partial protection of the region of the template strand in the vicinity of the startpoint, from –10 to +10. This suggests that $TF_{II}B$ is bound downstream of the TATA box, perhaps loosely associated with DNA and asymmetrically oriented with regard to the two DNA strands. The crystal structure shown in **Figure 21.16** extends this model. $TF_{II}B$ binds adjacent to TBP, extending contacts along one face of DNA. It makes contacts in the minor groove downstream of the TATA box, and contacts the major groove upstream of the TATA box, in a region called the BRE (2408). In archaea, the homologue of $TF_{II}B$ actually makes sequence-specific contacts with the promoter in the BRE region (652). $TF_{II}B$ may provide the surface that is in turn recognized by RNA polymerase, so that it is responsible for the directionality of the binding of the enzyme.

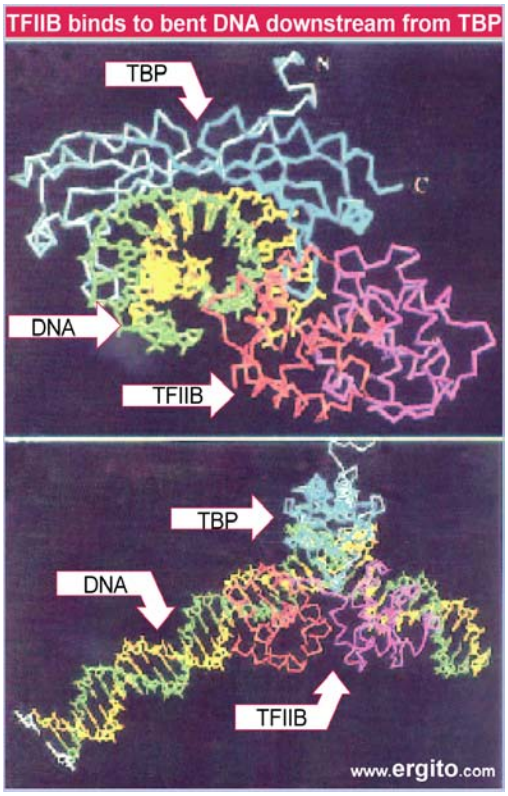

Figure 21.16 Two views of the ternary complex of TFIIB-TBP-DNA show that $TF_{II}B$ binds along the bent face of DNA. The two strands of DNA are green and yellow, TBP is blue, and TFIIB is red and purple. Photograph kindly provided by Stephen Burley.

The crystal structure of $TF_{II}B$ with RNA polymerase shows that three domains of the factor interact with the enzyme(4832). As illustrated schematically in **Figure 21.14**, an N-terminal zinc ribbon from $TF_{II}B$ contacts the enzyme near the site where RNA exits; it is possible that this interferes with the exit of RNA and influences the switch from abortive initiation to promoter escape. An elongated "finger" of $TF_{II}B$ is inserted into the polymerase active center. The C-terminal domain interacts with the RNA polymerase and with $TF_{II}D$ to orient the DNA. It also determines the path of the DNA where it contacts the factors $TF_{II}E$, $TF_{II}F$, and $TF_{II}H$, which may align them in the basal factor complex.

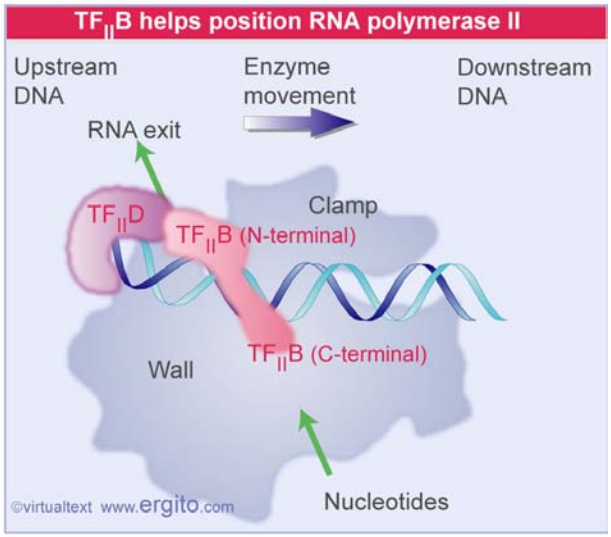

Figure 21.14 $TF_{II}B$ binds to DNA and contacts RNA polymerase near the RNA exit site and at the active center, and orients it on DNA. Compare with 1170, which shows the polymerase structure engaged in transcription.

The factor $TF_{II}F$ is a heterotetramer consisting of two types of subunit. The larger subunit (RAP74) has an ATP-dependent DNA helicase activity that could be involved in melting the DNA at initiation. The smaller subunit (RAP38) has some homology to the regions of bacterial sigma factor that contact the core polymerase; it binds tightly to RNA polymerase II. $TF_{II}F$ may bring RNA polymerase II to the assembling transcription complex and provide the means by which it binds. The complex of TBP and TAFs may interact with the CTD tail of RNA polymerase, and interaction with $TF_{II}B$ may also be important when $TF_{II}F$/polymerase joins the complex.

Polymerase binding extends the sites that are protected downstream to +15 on the template strand and +20 on the nontemplate strand. The enzyme extends the full length of the complex, since additional protection is seen at the upstream boundary.

What happens at TATA-less promoters? The same general transcription factors, including $TF_{II}D$, are needed. The Inr provides the positioning element; $TF_{II}D$ binds to it via an ability of one or more of the TAFs to recognize the Inr directly. Other TAFs in $TF_{II}D$ also recognize the DPE element downstream from the startpoint (3230). The function of TBP at these promoters is more like that at promoters for RNA polymerase I and at internal promoters for RNA polymerase III.

Assembly of the RNA polymerase II initiation complex provides an interesting contrast with prokaryotic transcription. Bacterial RNA polymerase is essentially a coherent aggregate with intrinsic ability to bind DNA; the sigma factor, needed for initiation but not for elongation, becomes part of the enzyme before DNA is bound, although it is later released. But RNA polymerase II can bind to the promoter only after separate transcription factors have bound. The factors play a role analogous to that of bacterial sigma factor – to allow the basic polymerase to recognize DNA

specifically at promoter sequences – but have evolved more independence. Indeed, the factors are primarily responsible for the specificity of promoter recognition. Only some of the factors participate in protein-DNA contacts (and only TBP makes sequence-specific contacts); thus protein-protein interactions are important in the assembly of the complex.

When a TATA box is present, it determines the location of the startpoint. Its deletion causes the site of initiation to become erratic, although any overall reduction in transcription is relatively small. Indeed, some TATA-less promoters lack unique startpoints; initiation occurs instead at any one of a cluster of startpoints. The TATA box aligns the RNA polymerase (via the interaction with $TF_{II}D$ and other factors) so that it initiates at the proper site. This explains why its location is fixed with respect to the startpoint. Binding of TBP to TATA is the predominant feature in recognition of the promoter, but two large TAFs ($TAF_{II}250$ and $TAF_{II}150$) also contact DNA in the vicinity of the startpoint and influence the efficiency of the reaction.

Although assembly can take place just at the core promoter *in vitro*, this reaction is not sufficient for transcription *in vivo*, where interactions with activators that recognize the more upstream elements are required. The activators interact with the basal apparatus at various stages during its assembly (see *Molecular Biology 5.22.5 Activators interact with the basal apparatus*).

Last updated on March 9, 2004

Reviews

223. Zawel, L. and Reinberg, D. (1993). *Initiation of transcription by RNA polymerase II: a multi-step process*. Prog. Nucleic Acid Res. Mol. Biol. 44, 67-108.

226. Nikolov, D. B. and Burley, S. K. (1997). *RNA polymerase II transcription initiation: a structural view*. Proc. Natl. Acad. Sci. USA 94, 15-22.

References

644. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). *Five intermediate complexes in transcription initiation by RNA polymerase II*. Cell 56, 549-561.

652. Nikolov, D. B. et al. (1995). *Crystal structure of a TFIIB-TBP-TATA-element ternary complex*. Nature 377, 119-128.

2408. Littlefield, O., Korkhin, Y., and Sigler, P. B. (1999). *The structural basis for the oriented assembly of a TBP/TFB/promoter complex.* Proc. Natl. Acad. Sci. USA 96, 13668-13673.

3230. Burke, T. W. and Kadonaga, J. T. (1996). *Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters.* Genes Dev. 10, 711-724.

4832. Bushnell, D. A., Westover, K. D., Davis, R. E., and Kornberg, R. D. (2004). *Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms.* Science 303, 983-988.

PROMOTERS AND ENHANCERS

5.21.11 Initiation is followed by promoter clearance

Key Concepts

- $TF_{II}E$ and $TF_{II}H$ are required to melt DNA to allow polymerase movement.
- Phosphorylation of the CTD may be required for elongation to begin.
- Further phosphorylation of the CTD is required at some promoters to end abortive initiation.
- The CTD may coordinate processing of RNA with transcription.

Most of the transcription factors are required solely to bind RNA polymerase to the promoter, but some act at a later stage (for review see 2407). Binding of $TF_{II}E$ causes the boundary of the region protected downstream to be extended by another turn of the double helix, to +30. Two further factors, $TF_{II}H$ and $TF_{II}J$, join the complex after $TF_{II}E$. They do not change the pattern of binding to DNA.

$TF_{II}H$ is the only general transcription factor that has independent enzymatic activities. Its several activities include an ATPase, helicases of both polarities, and a kinase activity that can phosphorylate the CTD tail of RNA polymerase II. $TF_{II}H$ is an exceptional factor that may play a role also in elongation. Its interaction with DNA downstream of the startpoint is required for RNA polymerase to escape from the promoter (2207). $TF_{II}H$ is also involved in repair of damage to DNA (see *Molecular Biology 5.21.12 A connection between transcription and repair*) (650).

The initiation reaction, as defined by formation of the first phosphodiester bond, occurs once RNA polymerase has bound. **Figure 21.17** proposes a model in which phosphorylation of the tail is needed to release RNA polymerase II from the transcription factors so that it can make the transition to the elongating form. Most of the transcription factors are released from the promoter at this stage.

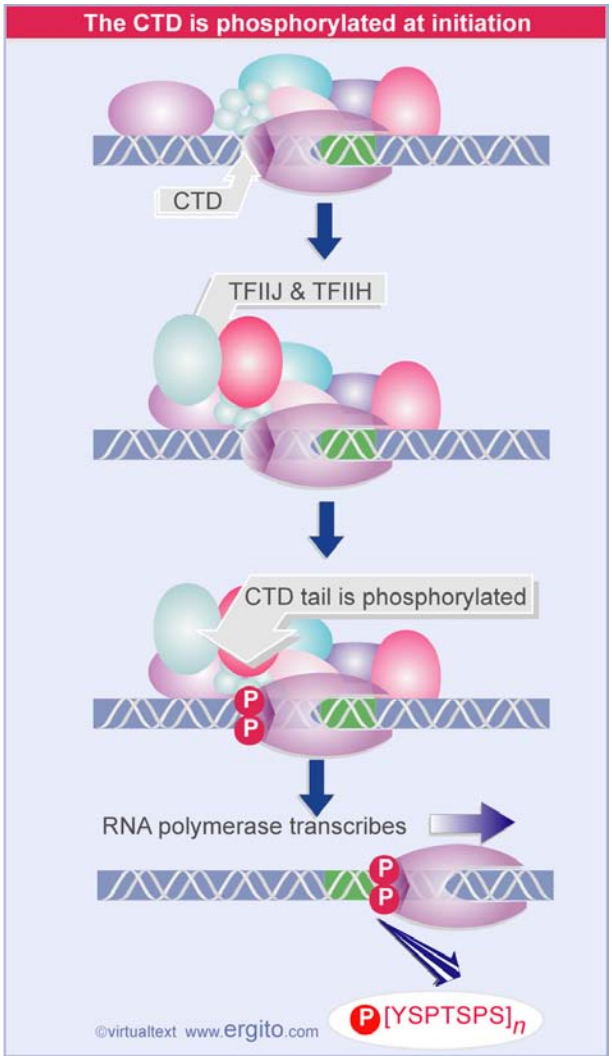

Figure 21.17 Phosphorylation of the CTD by the kinase activity of $TF_{II}H$ may be needed to release RNA polymerase to start transcription.

On a linear template, ATP hydrolysis, $TF_{II}E$, and the helicase activity of $TF_{II}H$ (provided by the XPB subunit) are required for polymerase movement. This requirement is bypassed with a supercoiled template. This suggests that $TF_{II}E$ and $TF_{II}H$ are required to melt DNA to allow polymerase movement to begin (946). The helicase activity of the XPB subunit of $TF_{II}H$ is responsible for the actual melting of DNA (2409; 2410).

RNA polymerase II stutters at some genes when it starts transcription. (The result is not dissimilar to the abortive initiation of bacterial RNA polymerase discussed in *Molecular Biology 3.9.11 Sigma factor controls binding to DNA*, although the mechanism is different.) At many genes, RNA polymerase II terminates after a short distance. The short RNA product is degraded rapidly. To extend elongation into the gene, a kinase called P-TEFb is required (for review see 948). This kinase is a member of the cdk family that controls the cell cycle (see *Molecular Biology 6.29 Cell cycle and growth regulation*). P-TEFb acts on the CTD, to

phosphorylate it further. We do not yet understand why this effect is required at some promoters but not others or how it is regulated.

The CTD may also be involved, directly or indirectly, in processing RNA after it has been synthesized by RNA polymerase II (for review see 2421, 4528). **Figure 21.18** summarizes processing reactions in which the CTD may be involved. The capping enzyme (guanylyl transferase), which adds the G residue to the 5′ end of newly synthesized mRNA, binds to the phosphorylated CTD: this may be important in enabling it to modify the 5′ end as soon as it is synthesized. A set of proteins called SCAFs bind to the CTD, and they may in turn bind to splicing factors. This may be a means of coordinating transcription and splicing. Some components of the cleavage/polyadenylation apparatus also bind to the CTD. Oddly enough, they do so at the time of initiation, so that RNA polymerase is all ready for the 3′ end processing reactions as soon as it sets out! All of this suggests that the CTD may be a general focus for connecting other processes with transcription (2239; for review see 2007, 4181). In the cases of capping and splicing, the CTD functions indirectly to promote formation of the protein complexes that undertake the reactions. In the case of 3′ end generation, it may participate directly in the reaction.

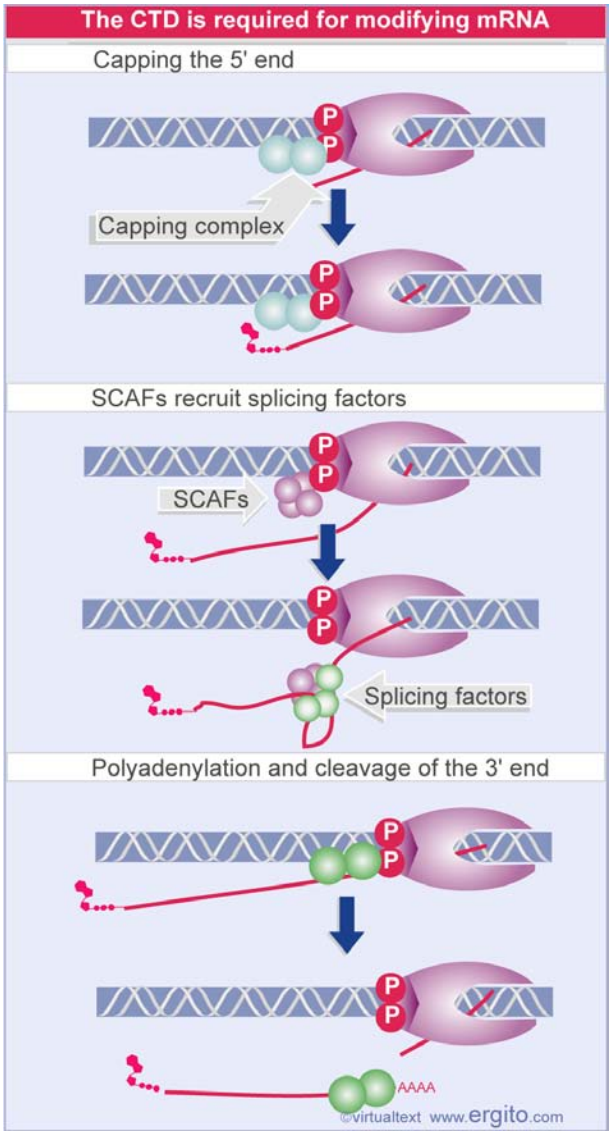

Figure 21.18 The CTD is important in recruiting enzymes that modify RNA.

The general process of initiation is similar to that catalyzed by bacterial RNA polymerase. Binding of RNA polymerase generates a closed complex, which is converted at a later stage to an open complex in which the DNA strands have been separated. In the bacterial reaction, formation of the open complex completes the necessary structural change to DNA; a difference in the eukaryotic reaction is that further unwinding of the template is needed after this stage.

Last updated on 9-24-2003

Reviews

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

2007. Hirose, Y. and Manley, J. L. (2000). *RNA polymerase II and the integration of nuclear events.* Genes Dev. 14, 1415-1429.

2407. Woychik, N. A. and Hampsey, M. (2002). *The RNA polymerase II machinery: structure illuminates function.* Cell 108, 453-463.

2421. Proudfoot, N. J., Furger, A., and Dye, M. J. (2002). *Integrating mRNA processing with transcription.* Cell 108, 501-512.

4181. Calvo, O. and Manley, J. L. (2003). *Strange bedfellows: polyadenylation factors at the promoter.* Genes Dev. 17, 1321-1327.

4528. Shilatifard, A., Conaway, R. C., and Conaway, J. W. (2003). *The RNA polymerase II elongation complex.* Annu. Rev. Biochem. 72, 693-715.

References

650. Goodrich, J. A. and Tjian, R. (1994). *Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II.* Cell 77, 145-156.

946. Holstege, F. C., van der Vliet, P. C., and Timmers, H. T. (1996). *Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH.* EMBO J. 15, 1666-1677.

2207. Spangler, L., Wang, X., Conaway, J. W., Conaway, R. C., and Dvir, A. (2001). *TFIIH action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA.* Proc. Natl. Acad. Sci. USA 98, 5544-5549.

2239. Fong, N. and Bentley, D. L. (2001). *Capping, splicing, and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD.* Genes Dev. 15, 1783-1795.

2409. Douziech, M., Coin, F., Chipoulet, J. M., Arai, Y., Ohkuma, Y., Egly, J. M., and Coulombe, B. (2000). *Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking.* Mol. Cell Biol. 20, 8168-8177.

2410. Kim, T. K., Ebright, R. H., and Reinberg, D. (2000). *Mechanism of ATP-dependent promoter melting by transcription factor IIH.* Science 288, 1418-1422.

PROMOTERS AND ENHANCERS

5.21.12 A connection between transcription and repair

Key Concepts

- Transcribed genes are preferentially repaired when DNA damage occurs.
- $TF_{II}H$ provides the link to a complex of repair enzymes.
- Mutations in the XPD component of $TF_{II}H$ cause three types of human diseases

In both bacteria and eukaryotes, there is a direct link from RNA polymerase to the activation of repair. The basic phenomenon was first observed because transcribed genes are preferentially repaired. Then it was discovered that it is only the template strand of DNA that is the target – the nontemplate strand is repaired at the same rate as bulk DNA.

In bacteria, the repair activity is provided by the *uvr* excision-repair system (see *Molecular Biology 4.15.21 Excision repair systems in E. coli*). Preferential repair is abolished by mutations in the gene *mfd*, whose product provides the link from RNA polymerase to the Uvr enzymes (for review see 224).

Figure 21.19 shows a model for the link between transcription and repair. When RNA polymerase encounters DNA damage in the template strand, it stalls because it cannot use the damaged sequences as a template to direct complementary base pairing. This explains the specificity of the effect for the template strand (damage in the nontemplate strand does not impede progress of the RNA polymerase).

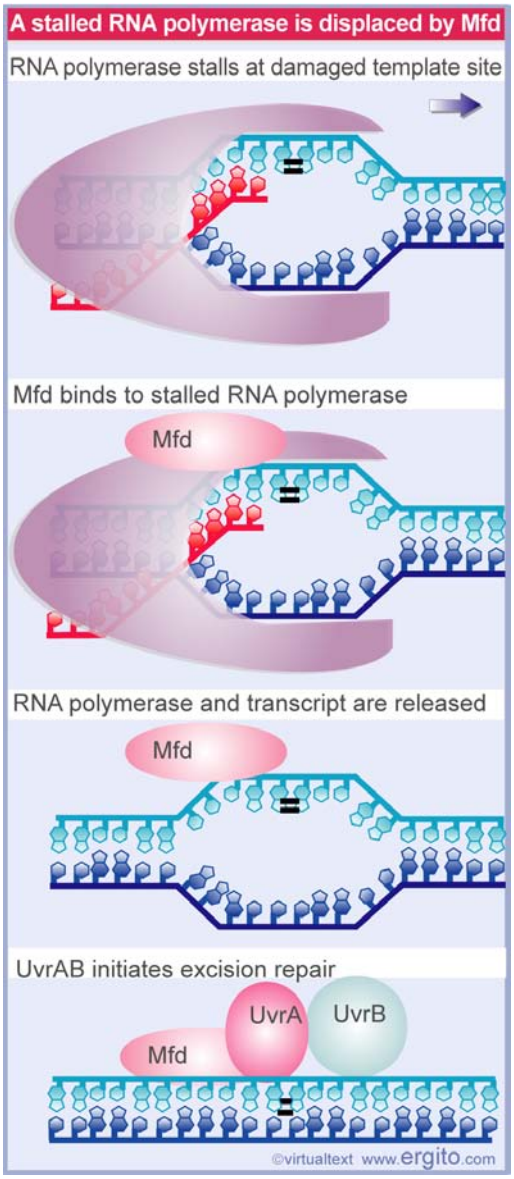

Figure 21.19 Mfd recognizes a stalled RNA polymerase and directs DNA repair to the damaged template strand.

The Mfd protein has two roles. First, it displaces the ternary complex of RNA polymerase from DNA. Second, it causes the UvrABC enzyme to bind to the damaged DNA. This leads to repair of DNA by the excision-repair mechanism (see **Figure 15.40**). After the DNA has been repaired, the next RNA polymerase to traverse the gene is able to produce a normal transcript (661).

A similar mechanism, although relying on different components, is used in eukaryotes. The template strand of a transcribed gene is preferentially repaired following UV-induced damage. The general transcription factor $TF_{II}H$ is involved. $TF_{II}H$ is found in alternative forms, which consist of a core associated with other subunits.

$TF_{II}H$ has a common function in both initiating transcription and repairing damage. The same helicase subunit (XPD) creates the initial transcription bubble and melts DNA at a damaged site. Its other functions differ between transcription and repair, as provided by the appropriate form of the complex.

Figure 21.20 shows that the basic factor involved in transcription consists of a core (of 5 subunits) associated with other subunits that have a kinase activity; this complex also includes a repair subunit. The kinase catalytic subunit that phosphorylates the CTD of RNA polymerase belongs to a group of kinases that are involved in cell cycle control (see *Molecular Biology 6.29 Cell cycle and growth regulation*). It is possible that this connection influences transcription in response to the stage of the cell cycle.

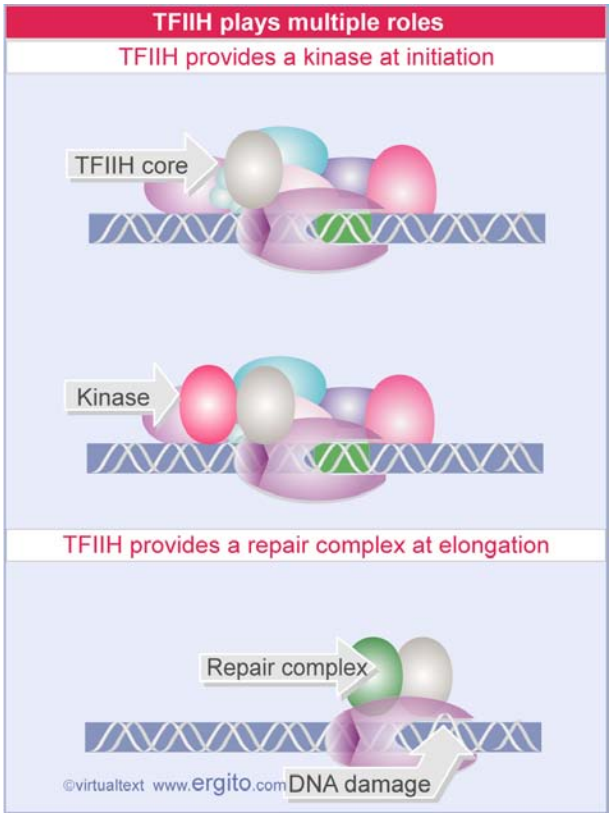

Figure 21.20 The $TF_{II}H$ core may associate with a kinase at initiation and associate with a repair complex when damaged DNA is encountered.

The alternative complex consists of the core associated with a large group of proteins that are coded by repair genes. (The basic model for repair is shown in **Figure 15.53**.) The repair proteins include a subunit (XPC) that recognizes damaged DNA, which provides the coupling function that enables a template strand to be preferentially repaired when RNA polymerase becomes stalled at damaged DNA. Other proteins associated with the complex include endonucleases (XPG, XPF, ERCC1). Homologous proteins are found in the complexes in yeast (where they are often identified by *rad* mutations that are defective in repair) and in Man (where they are identified by mutations that cause diseases resulting from deficiencies in

repairing damaged DNA) (662; 663). (Subunits with the name XP are coded by genes in which mutations cause the disease xeroderma pigmentosum (see *Molecular Biology 4.15.28 Eukaryotic cells have conserved repair systems*).

The kinase complex and the repair complex can associate and dissociate reversibly from the core $TF_{II}H$. This suggests a model in which the first form of $TF_{II}H$ is required for initiation, but may be replaced by the other form (perhaps in response to encountering DNA damage). $TF_{II}H$ dissociates from RNA polymerase at an early stage of elongation (after transcription of ~50 bp); its reassociation at a site of damaged DNA may require additional coupling components.

The repair function may require modification or degradation of RNA polymerase. The large subunit of RNA polymerase is degraded when the enzyme stalls at sites of UV damage. We do not yet understand the connection between the transcription/repair apparatus as such and the degradation of RNA polymerase. It is possible that removal of the polymerase is necessary once it has become stalled (664).

This degradation of RNA polymerase is deficient in cells from patients with Cockayne's syndrome (a repair disorder). Cockayne's syndrome is caused by mutations in either of two genes (*CSA* and *CSB*), both of whose products appear to be part of or bound to $TF_{II}H$. Cockayne's syndrome is also occasionally caused by mutations in XPD.

Another disease that can be caused by mutations in XPD is trichothiodystrophy, which has little in common with XP or Cockayne's (it involves mental retardation and is marked by changes in the structure of hair). All of this marks XPD as a pleiotropic protein, in which different mutations can affect different functions. In fact, XPD is required for the stability of the $TF_{II}H$ complex during transcription, but the helicase activity as such is not needed. Mutations that prevent XPD from stabilizing the complex cause trichothiodystrophy. The helicase activity is required for the repair function. Mutations that affect the helicase activity cause the repair deficiency that results in XP or Cockayne's syndrome (for review see 1641).

Last updated on 4-30-2001

Reviews

224. Selby, C. P. and Sancar, A. (1994). *Mechanisms of transcription-repair coupling and mutation frequency decline*. Microbiol. Rev. 58, 317-329.

1641. Lehmann, A. R. (2001). *The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases.* Genes Dev. 15, 15-23.

References

661. Selby, C. P. and Sancar, A. (1993). *Molecular mechanism of transcription-repair coupling*. Science 260, 53-58.

662. Schaeffer, L. et al. (1993). *DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor*. Science 260, 58-63.

663. Svejstrup, J. Q. et al. (1995). *Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome*. Cell 80, 21-28.

664. Bregman, D. et al. (1996). *UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells*. Proc. Natl. Acad. Sci. USA 93, 11586-11590.

PROMOTERS AND ENHANCERS

5.21.13 Short sequence elements bind activators

Key Terms

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 **CAAT box** is part of a conserved sequence located upstream of the startpoints of eukaryotic transcription units; it is recognized by a large group of transcription factors.

The **GC box** is a common pol II promoter element consisting of the sequence GGGCGG.

Key Concepts

- Short conserved sequence elements are dispersed in the region preceding the startpoint.
- The upstream elements increase the frequency of initiation.
- The factors that bind to them to stimulate transcription are called activators.

A promoter for RNA polymerase II consists of two types of region. The startpoint itself is identified by the Inr and/or by the TATA box close by. In conjunction with the general transcription factors, RNA polymerase II forms an initiation complex surrounding the startpoint, as we have just described. The efficiency and specificity with which a promoter is recognized, however, depend upon short sequences, farther upstream, which are recognized by a different group of factors, usually called **activators**. Usually the target sequences are ~100 bp upstream of the startpoint, but sometimes they are more distant. Binding of activators at these sites may influence the formation of the initiation complex at (probably) any one of several stages.

An analysis of a typical promoter is summarized in **Figure 21.21**. Individual base substitutions were introduced at almost every position in the 100 bp upstream of the β-globin startpoint. The striking result is that *most mutations do not affect the ability of the promoter to initiate transcription.* Down mutations occur in three locations, corresponding to three short discrete elements. The two upstream elements have a greater effect on the level of transcription than the element closest to the startpoint. Up mutations occur in only one of the elements. We conclude that the three short sequences centered at –30, –75, and –90 constitute the promoter. Each of them corresponds to the consensus sequence for a common type of promoter element.

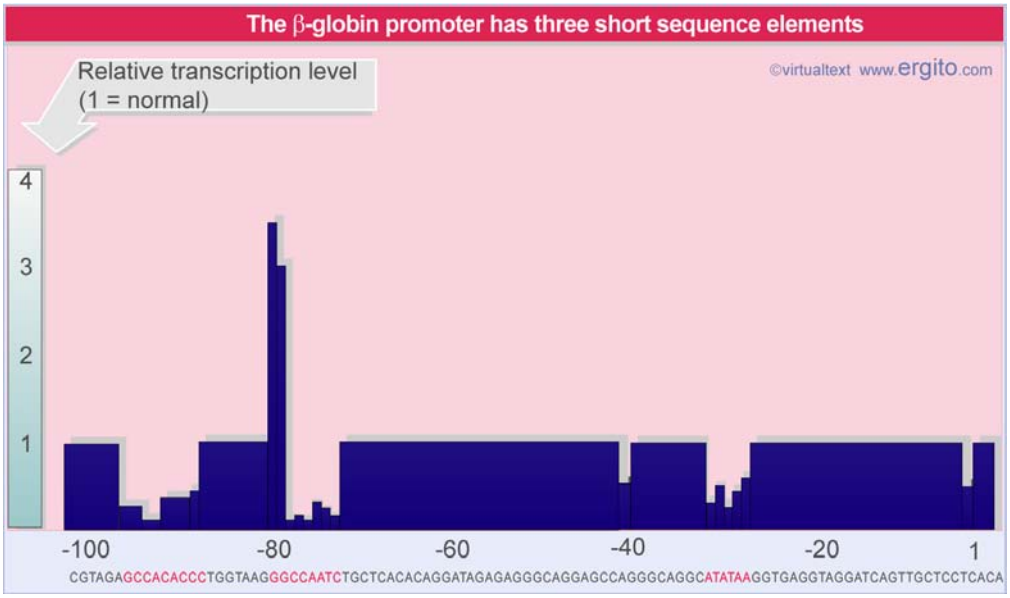

Figure 21.21 Saturation mutagenesis of the upstream region of the β -globin promoter identifies three short regions (centered at -30, -75, and -90) that are needed to initiate transcription. These correspond to the TATA, CAAT, and GC boxes.

The TATA box (centered at –30) is the least effective component of the promoter as measured by the reduction in transcription that is caused by mutations. But although initiation is not prevented when a TATA box is mutated, the startpoint varies from its usual precise location. This confirms the role of the TATA box as a crucial positioning component of the core promoter.

The basal elements and the elements upstream of them have different types of functions. The basal elements (the TATA box and Inr) primarily determine the location of the startpoint, but can sponsor initiation only at a rather low level. They identify the *location* at which the general transcription factors assemble to form the basal complex. The sequence elements farther upstream influence the *frequency* of initiation, most likely by acting directly on the general transcription factors to enhance the efficiency of assembly into an initiation complex (see *Molecular Biology 5.22.5 Activators interact with the basal apparatus*).

The sequence at –75 is the **CAAT box**. Named for its consensus sequence, it was one of the first common elements to be described. It is often located close to –80, but it can function at distances that vary considerably from the startpoint. It functions in either orientation. Susceptibility to mutations suggests that the CAAT box plays a strong role in determining the efficiency of the promoter, but does not influence its specificity.

The **GC box** at –90 contains the sequence GGGCGG. Often multiple copies are present in the promoter, and they occur in either orientation. It too is a relatively common promoter component.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.21.13

PROMOTERS AND ENHANCERS

5.21.14 Promoter construction is flexible but context can be important

Key Concepts

- No individual upstream element is essential for promoter function, although one or more elements must be present for efficient initiation.
- Some elements are recognized by multiple factors, and the factor that is used at any particular promoter may be determined by the context of the other factors that are bound.

Promoters are organized on a principle of "mix and match." A variety of elements can contribute to promoter function, but none is essential for all promoters. Some examples are summarized in **Figure 21.22**. Four types of elements are found altogether in these promoters: TATA, GC boxes, CAAT boxes, and the octamer (an 8 bp element). The elements found in any individual promoter differ in number, location, and orientation. No element is common to all of the promoters. Although the promoter conveys directional information (transcription proceeds only in the downstream direction), the GC and CAAT boxes seem to be able to function in either orientation. This implies that the elements function solely as DNA-binding sites to bring transcription factors into the vicinity of the startpoint; the structure of a factor must be flexible enough to allow it to make protein-protein contacts with the basal apparatus irrespective of the way in which its DNA-binding domain is oriented and its exact distance from the startpoint.

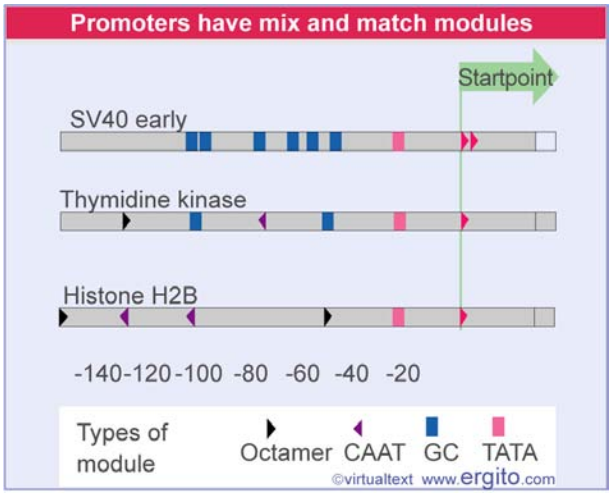

Figure 21.22 Promoters contain different combinations of TATA boxes, CAAT boxes, GC boxes, and other elements.

Activators that are more or less ubiquitous are assumed to be available to any promoter that has a copy of the element that they recognize. Common elements

recognized by ubiquitous activators include the CAAT box, GC box, and the octamer. All promoters probably require one or more of these elements in order to function efficiently. An activator typically has a consensus sequence of <10 bp, but actually covers a length of ~20 bp of DNA. Given the sizes of the activators, and the length of DNA each covers, we expect that the various proteins will together cover the entire region upstream of the startpoint in which the elements reside.

Usually a particular consensus sequence is recognized by a corresponding activator (or by a member of a family of factors). However, sometimes a particular promoter sequence can be recognized by one of several activators. A ubiquitous activator, Oct-1, binds to the octamer to activate the histone H2B (and presumably also other) genes. Oct-1 is the only octamer-binding factor in nonlymphoid cells. But in lymphoid cells, a different activator, Oct-2, binds to the octamer to activate the immunoglobulin κ light gene. So Oct-2 is a tissue-specific activator, while Oct-1 is ubiquitous. The exact details of recognition are not so important to know as the fact that a variety of activators recognize CAAT boxes.

The use of the same octamer in the ubiquitously expressed H2B gene and the lymphoid-specific immunoglobulin genes poses a paradox. Why does the ubiquitous Oct-1 fail to activate the immunoglobulin genes in nonlymphoid tissues? The *context* must be important: Oct-2 rather than Oct-1 may be needed to interact with other proteins that bind at the promoter. These results mean that we cannot predict whether a gene will be activated by a particular activator simply on the basis of the presence of particular elements in its promoter.

PROMOTERS AND ENHANCERS

5.21.15 Enhancers contain bidirectional elements that assist initiation

Key Terms

An **enhancer** is a *cis*-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter.

An **upstream activator sequence (UAS)** is the equivalent in yeast of the enhancer in higher eukaryotes.

Key Concepts

- An enhancer activates the nearest promoter to it, and can be any distance either upstream or downstream of the promoter.
- A UAS (upstream activator sequence) in yeast behaves like an enhancer but works only upstream of the promoter.
- Similar sequence elements are found in enhancers and promoters.
- Enhancers form complexes of activators that interact directly or indirectly with the promoter.

We have considered the promoter so far as an isolated region responsible for binding RNA polymerase. But eukaryotic promoters do not necessarily function alone. In at least some cases, the activity of a promoter is enormously increased by the presence of an **enhancer**, which consists of another group of elements, but located at a variable distance from those regarded as comprising part of the promoter itself (665; for review see 219).

The concept that the enhancer is distinct from the promoter reflects two characteristics. The position of the enhancer relative to the promoter need not be fixed, but can vary substantially. **Figure 21.23** shows that it can be either upstream or downstream. And it can function in either orientation (that is, it can be inverted) relative to the promoter. Manipulations of DNA show that an enhancer can stimulate any promoter placed in its vicinity. In natural genomes, enhancers can be located within genes (that is, just downstream of the promoter) or tens of kilobases away in either direction.

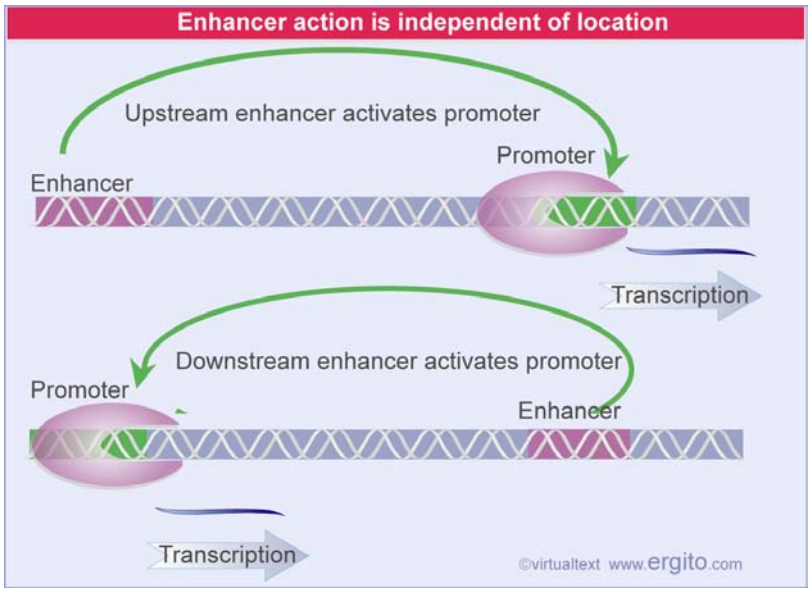

Figure 21.23 An enhancer can activate a promoter from upstream or downstream locations, and its sequence can be inverted relative to the promoter,

For operational purposes, it is sometimes useful to define the promoter *as a sequence or sequences of DNA that must be in a (relatively) fixed location with regard to the startpoint.* By this definition, the TATA box and other upstream elements are included, but the enhancer is excluded. This is, however, a working definition rather than a rigid classification.

Elements analogous to enhancers, called upstream activator sequences (**UAS**), are found in yeast. They can function in either orientation, at variable distances upstream of the promoter, but cannot function when located downstream. They have a regulatory role: in several cases the UAS is bound by the regulatory protein(s) that activates the genes downstream.

Reconstruction experiments in which the enhancer sequence is removed from the DNA and then is inserted elsewhere show that normal transcription can be sustained so long as it is present *anywhere* on the DNA molecule. If a β-globin gene is placed on a DNA molecule that contains an enhancer, its transcription is increased *in vivo* more than 200-fold, even when the enhancer is several kb upstream or downstream of the startpoint, in either orientation. We have yet to discover at what distance the enhancer fails to work.

Reviews

219. Muller, M. M., Gerster, T., and Schaffner, W. (1988). *Enhancer sequences and the regulation of gene transcription.* Eur. J. Biochem. 176, 485-495.

References

665. Banerji, J., Rusconi, S., and Schaffner, W. (1981). *Expression of β -globin gene is enhanced by remote SV40 DNA sequences*. Cell 27, 299-308.

PROMOTERS AND ENHANCERS

5.21.16 Enhancers contain the same elements that are found at promoters

Key Terms

An **enhanceosome** is a complex of transcription factors that assembles cooperatively at an enhancer.

Key Concepts

- Enhancers are made of the same short sequence elements that are found in promoters.
- The density of sequence components is greater in the enhancer than in the promoter.

A difference between the enhancer and a typical promoter is presented by the density of regulatory elements. **Figure 21.24** summarizes the susceptibility of the SV40 enhancer to damage by mutation; and we see that a much greater proportion of its sites directly influences its function than is the case with the promoter analyzed in the same way in **Figure 21.21**. There is a corresponding increase in the density of protein-binding sites. Many of these sites are common elements in promoters; for example, AP1 and the octamer.

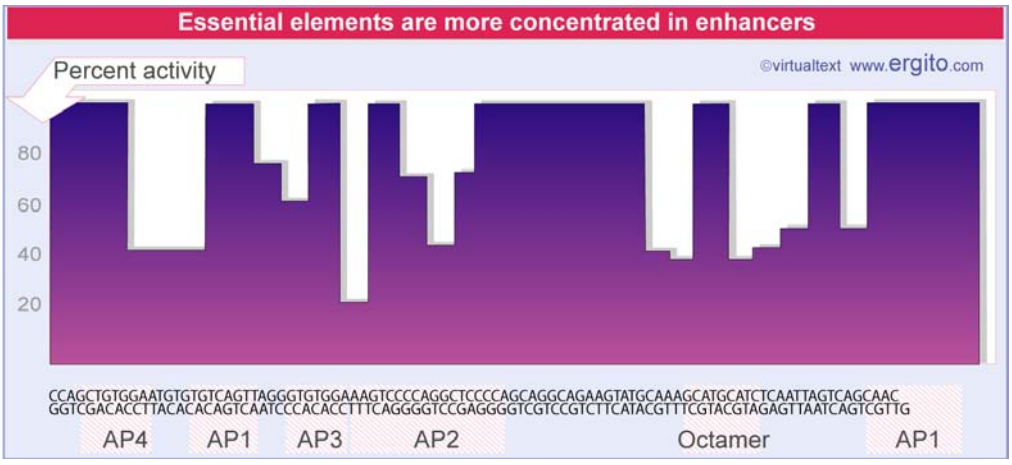

Figure 21.24 An enhancer contains several structural motifs. The histogram plots the effect of all mutations that reduce enhancer function to <75% of wild type. Binding sites for proteins are indicated below the histogram.

The specificity of transcription may be controlled by either a promoter or an enhancer. A promoter may be specifically regulated, and a nearby enhancer used to increase the efficiency of initiation; or a promoter may lack specific regulation, but become active only when a nearby enhancer is specifically activated. An example is provided by immunoglobulin genes, which carry enhancers *within* the transcription

unit. The immunoglobulin enhancers appear to be active only in the B lymphocytes in which the immunoglobulin genes are expressed. Such enhancers provide part of the regulatory network by which gene expression is controlled.

A difference between enhancers and promoters may be that an enhancer shows greater cooperativity between the binding of factors. A complex that assembles at the enhancer that responds to IFN (interferon) γ assembles cooperatively to form a functional structure called the **enhanceosome**. Binding of the nonhistone protein HMGI(Y) bends the DNA into a structure that then binds several activators (NF- κB, IRF, ATF-Jun). In contrast with the "mix and match" construction of promoters, all of these components are required to create an active structure at the enhancer. These components do not themselves directly bind to RNA polymerase, but they create a surface that binds a *coactivating complex*. The complex helps the pre-initiation complex of basal transcription factors that is assembling at the promoter to recruit RNA polymerase (for review see 1866; 1867). We discuss the function of coactivators in more detail in *Molecular Biology 5.22.5 Activators interact with the basal apparatus*.

Reviews

1866. Maniatis, T., Falvo, J. V., Kim, T. H., Kim, T. K., Lin, C. H., Parekh, B. S., and Wathelet, M. G. (1998). *Structure and function of the interferon-beta enhanceosome.* Cold Spring Harbor Symp. Quant. Biol. 63, 609-620.

1867. Munshi, N., Yie, Y., Merika, M., Senger, K., Lomvardas, S., Agalioti, T., and Thanos, D. (1999). *The IFN-beta enhancer: a paradigm for understanding activation and repression of inducible gene expression.* Cold Spring Harbor Symp. Quant. Biol. 64, 149-159.

PROMOTERS AND ENHANCERS

5.21.17 Enhancers work by increasing the concentration of activators near the promoter

Key Concepts

- Enhancers usually work only in *cis* configuration with a target promoter.
- Enhancers can be made to work in *trans* configuration by linking the DNA that contains the target promoter to the DNA that contains the enhancer via a protein bridge or by catenating the two molecules.
- The principle is that an enhancer works in any situation in which it is constrained to be in the same proximity as the promoter.

How can an enhancer stimulate initiation at a promoter that can be located any distance away on either side of it? When enhancers were first discovered, several possibilities were considered for their action as elements distinctly different from promoters (for review see 1715):

- An enhancer could change the overall structure of the template – for example, by influencing the density of supercoiling.
- It could be responsible for locating the template at a particular place within the cell – for example, attaching it to the nuclear matrix.
- An enhancer could provide an "entry site," a point at which RNA polymerase (or some other essential protein) initially associates with chromatin.

Now we take the view that enhancer function involves the same sort of interaction with the basal apparatus as the interactions sponsored by upstream promoter elements. Enhancers are modular, like promoters. Some elements are found in both enhancers and promoters. Some individual elements found in promoters share with enhancers the ability to function at variable distance and in either orientation. So the distinction between enhancers and promoters is blurred: enhancers might be viewed as containing promoter elements that are grouped closely together, with the ability to function at increased distances from the startpoint (666).

The essential role of the enhancer may be to increase the concentration of activator in the vicinity of the promoter (vicinity in this sense being a relative term). Two types of experiment illustrated in **Figure 21.25** suggest that this is the case.

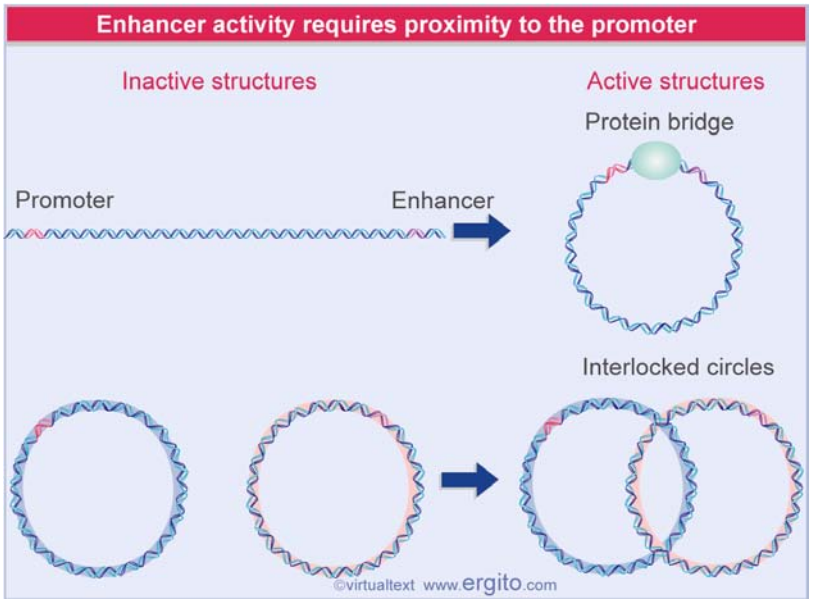

Figure 21.25 An enhancer may function by bringing proteins into the vicinity of the promoter. An enhancer does not act on a promoter at the opposite end of a long linear DNA, but becomes effective when the DNA is joined into a circle by a protein bridge. An enhancer and promoter on separate circular DNAs do not interact, but can interact when the two molecules are catenated.

A fragment of DNA that contains an enhancer at one end and a promoter at the other is not effectively transcribed, but the enhancer can stimulate transcription from the promoter when they are connected by a protein bridge. Since structural effects, such as changes in supercoiling, could not be transmitted across such a bridge, this suggests that the critical feature is bringing the enhancer and promoter into close proximity (667).

A bacterial enhancer provides a binding site for the regulator NtrC, which acts upon RNA polymerase using promoters recognized by σ^{54}. When the enhancer is placed upon a circle of DNA that is catenated (interlocked) with a circle that contains the promoter, initiation is almost as effective as when the enhancer and promoter are on the same circular molecule. But there is no initiation when the enhancer and promoter are on separated circles. Again this suggests that the critical feature is localization of the protein bound at the enhancer, to increase its chance of contacting a protein bound at the promoter.

If proteins bound at an enhancer several kb distant from a promoter interact directly with proteins bound in the vicinity of the startpoint, the organization of DNA must be flexible enough to allow the enhancer and promoter to be closely located. This requires the intervening DNA to be extruded as a large "loop." Such loops have been directly observed in the case of the bacterial enhancer.

There is an interesting exception to the rule that enhancers are *cis*-acting in natural situations. This is seen in the phenomenon of transvection. Pairing of somatic chromosomes allows an enhancer on one chromosome to activate a promoter on the partner chromosome. This reinforces the view that enhancers work by proximity.

What limits the activity of an enhancer? Typically it works upon the nearest promoter. There are situations in which an enhancer is located between two promoters, but activates only one of them on the basis of specific protein-protein contacts between the complexes bound at the two elements. The action of an enhancer may be limited by an insulator – an element in DNA that prevents it from acting on promoters beyond (see *Molecular Biology 5.21.20 Insulators block the actions of enhancers and heterochromatin*).

The generality of enhancement is not yet clear. We do not know what proportion of cellular promoters require an enhancer to achieve their usual level of expression. Nor do we know how often an enhancer provides a target for regulation. Some enhancers are activated only in the tissues in which their genes function, but others could be active in all cells.

Last updated on February 4, 2004

Reviews

1715. Blackwood, E. M. and Kadonaga, J. T. (1998). *Going the distance: a current view of enhancer action.* Science 281, 60-63.

References

666. Zenke, M. et al. (1986). *Multiple sequence motifs are involved in SV40 enhancer function*. EMBO J. 5, 387-397.

667. Mueller-Storm, H. P., Sogo, J. M., and Schaffner, W. (1989). *An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge*. Cell 58, 767-777.

PROMOTERS AND ENHANCERS

5.21.18 Gene expression is associated with demethylation

Key Terms

Isoschizomers are restriction enzymes that cleave the same DNA sequence but are affected differently by its state of methylation.

Key Concepts

- Demethylation at the 5′ end of the gene is necessary for transcription.

Methylation of DNA is one of the parameters that controls transcription. Methylation in the vicinity of the promoter is associated with the absence of transcription. This is one of several regulatory events that influence the activity of a promoter; like the other regulatory events, typically this will apply to both (allelic) copies of the gene. However, methylation also occurs as an epigenetic event that can distinguish alleles whose sequences are identical. This can result in differences in the expression of the paternal and maternal alleles (see *Molecular Biology 5.23.20 DNA methylation is responsible for imprinting*). In this chapter we are concerned with the means by which methylation influences transcription.

The distribution of methyl groups can be examined by taking advantage of restriction enzymes that cleave target sites containing the CG doublet. Two types of restriction activity are compared in **Figure 21.26**. These **isoschizomers** are enzymes that cleave the same target sequence in DNA, but have a different response to its state of methylation.

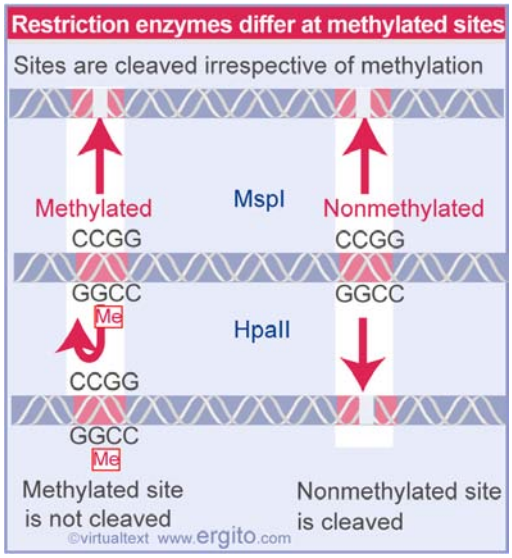

Figure 21.26 The restriction enzyme MspI cleaves all CCGG sequences whether or not they are methylated at the second C, but HpaII cleaves only nonmethylated CCGG tetramers.

The enzyme HpaII cleaves the sequence CCGG (writing the sequence of only one strand of DNA). But if the second C is methylated, the enzyme can no longer recognize the site. However, the enzyme MspI cleaves the same target site *irrespective* of the state of methylation at this C. So MspI can be used to identify all the CCGG sequences; and HpaII can be used to determine whether or not they are methylated.

With a substrate of nonmethylated DNA, the two enzymes would generate the same restriction bands. But in methylated DNA, the modified positions are not cleaved by HpaII. For every such position, one larger HpaII fragment replaces two MspI fragments. **Figure 21.27** gives an example.

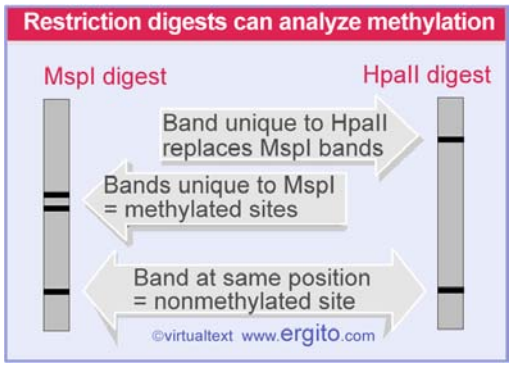

Figure 21.27 The results of MspI and HpaII cleavage are compared by gel electrophoresis of the fragments.

Many genes show a pattern in which the state of methylation is constant at most sites, but varies at others. Some of the sites are methylated in all tissues examined;

some sites are unmethylated in all tissues. *A minority of sites are methylated in tissues in which the gene is not expressed, but are not methylated in tissues in which the gene is active.* So an active gene may be described as *undermethylated.*

Experiments with the drug 5-azacytidine produce indirect evidence that demethylation can result in gene expression. The drug is incorporated into DNA in place of cytidine, and cannot be methylated, because the 5′ position is blocked. This leads to the appearance of demethylated sites in DNA as the consequence of replication (following the scheme on the right of **Figure 14.35**).

The phenotypic effects of 5-azacytidine include the induction of changes in the state of cellular differentiation; for example, muscle cells are induced to develop from nonmuscle cell precursors. The drug also activates genes on a silent X chromosome, which raises the possibility that the state of methylation could be connected with chromosomal inactivity.

As well as examining the state of methylation of resident genes, we can compare the results of introducing methylated or nonmethylated DNA into new host cells. Such experiments show a clear correlation: *the methylated gene is inactive, but the nonmethylated gene is active.*

What is the extent of the undermethylated region? In the chicken α-globin gene cluster in adult erythroid cells, the undermethylation is confined to sites that extend from ~500 bp upstream of the first of the two adult α genes to ~500 bp downstream of the second. Sites of undermethylation are present in the entire region, including the spacer between the genes. The region of undermethylation coincides with the region of maximum sensitivity to DNAase I. This argues that undermethylation is a feature of a domain that contains a transcribed gene or genes. As with other changes in chromatin, it seems likely that the absence of methyl groups is associated with the *ability to be transcribed* rather than with the act of transcription itself.

Our problem in interpreting the general association between undermethylation and gene activation is that only a minority (sometimes a small minority) of the methylated sites are involved. It is likely that the state of methylation is critical at specific sites or in a restricted region. It is also possible that a reduction in the level of methylation (or even the complete removal of methyl groups from some stretch of DNA) is part of some structural change needed to permit transcription to proceed.

In particular, demethylation at the promoter may be necessary to make it available for the initiation of transcription. In the γ-globin gene, for example, the presence of methyl groups in the region around the startpoint, between –200 and +90, suppresses transcription. Removal of the 3 methyl groups located upstream of the startpoint or of the 3 methyl groups located downstream does not relieve the suppression. But removal of all methyl groups allows the promoter to function. Transcription may therefore require a methyl-free region at the promoter (see *Molecular Biology 5.21.19 CpG islands are regulatory targets*). There are exceptions to this general relationship.

Some genes can be expressed even when they are extensively methylated. Any connection between methylation and expression thus is not universal in an organism, but the general rule is that methylation prevents gene expression and demethylation

is required for expression.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.21.18

PROMOTERS AND ENHANCERS

5.21.19 CpG islands are regulatory targets

Key Terms

CpG island is a stretch of 1-2 kb in a mammalian genome that is rich in unmethylated CpG doublets.

Key Concepts

- CpG islands surround the promoters of constitutively expressed genes where they are unmethylated.
- They are also found at the promoters of some tissue-regulated genes.
- There are ~29,000 CpG islands in the human genome.
- Methylation of a CpG island prevents activation of a promoter within it.
- Repression is caused by proteins that bind to methylated CpG doublets.

The presence of **CpG islands** in the 5′ regions of some genes is connected with the effect of methylation on gene expression. These islands are detected by the presence of an increased density of the dinucleotide sequence, CpG.

The CpG doublet occurs in vertebrate DNA at only ~20% of the frequency that would be expected from the proportion of G·C base pairs. (This may be because CpG doublets are methylated on C, and spontaneous deamination of methyl-C converts it to T, introducing a mutation that removes the doublet.) In certain regions, however, the density of CpG doublets reaches the predicted value; in fact, it is increased by 10× relative to the rest of the genome. The CpG doublets in these regions are unmethylated.

These CpG-rich islands have an average G·C content of ~60%, compared with the 40% average in bulk DNA. They take the form of stretches of DNA typically 1-2 kb long. There are ~45,000 such islands altogether in the human genome (1443). Some of the islands are present in repeated Alu elements, and may just be the consequence of their high G·C-content. The human genome sequence confirms that, excluding these, there are ~29,000 islands. There are fewer in the mouse genome, ~15,500. About 10,000 of the predicted islands in both species appear to reside in a context of sequences that are conserved between the species, suggesting that these may be the islands with regulatory significance. The structure of chromatin in these regions has changes associated with gene expression (see *Molecular Biology 5.23.11 Promoter activation involves an ordered series of events*); there is a reduced content of histone H1 (which probably means that the structure is less compact), the other histones are extensively acetylated (a feature that tends to be associated with gene expression), and there are hypersensitive sites (as would be expected of active promoters) (710).

In several cases, CpG-rich islands begin just upstream of a promoter and extend

downstream into the transcribed region before petering out. **Figure 21.28** compares the density of CpG doublets in a "general" region of the genome with a CpG island identified from the DNA sequence. The CpG island surrounds the 5′ region of the APRT gene, which is constitutively expressed.

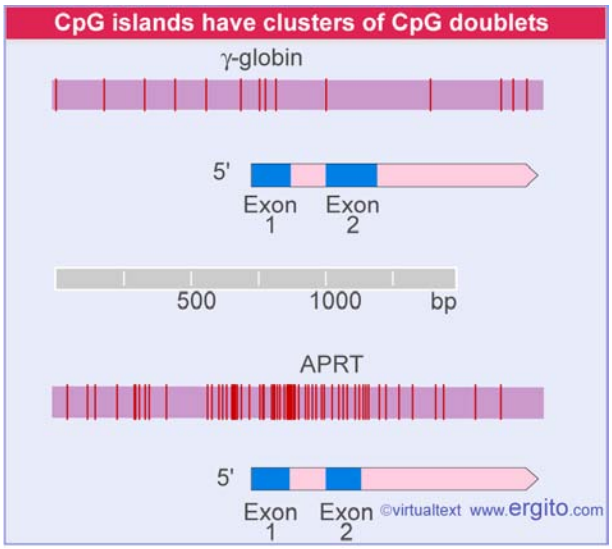

Figure 21.28 The typical density of CpG doublets in mammalian DNA is ~1/100 bp, as seen for a γ -globin gene. In a CpG-rich island, the density is increased to >10 doublets/100 bp. The island in the APRT gene starts ~100 bp upstream of the promoter and extends ~400 bp into the gene. Each vertical line represents a CpG doublet.

All of the "housekeeping" genes that are constitutively expressed have CpG islands; this accounts for about half of the islands altogether. The other half of the islands occur at the promoters of tissue-regulated genes; only a minority (<40%) of these genes have islands. In these cases, the islands are unmethylated irrespective of the state of expression of the gene. The presence of unmethylated CpG-rich islands may be necessary, but therefore is not sufficient, for transcription. So the presence of unmethylated CpG islands may be taken as an indication that a gene is potentially active, rather than inevitably transcribed. Many islands that are nonmethylated in the animal become methylated in cell lines in tissue culture, and this could be connected with the inability of these lines to express all of the functions typical of the tissue from which they were derived.

Methylation of a CpG island can affect transcription. Two mechanisms can be involved:

- Methylation of a binding site for some factor may prevent it from binding. This happens in a case of binding to a regulatory site other than the promoter (see *Molecular Biology 5.23.21 Oppositely imprinted genes can be controlled by a single center*).
- Or methylation may cause specific repressors to bind to the DNA.

Repression is caused by either of two types of protein that bind to methylated CpG sequences (for review see 2424). The protein MeCP1 requires the presence of several methyl groups to bind to DNA, while MeCP2 and a family of related proteins can bind to a single methylated CpG base pair. This explains why a methylation-free zone is required for initiation of transcription. Binding of proteins of either type prevents transcription *in vitro* by a nuclear extract (711).

MeCP2, which directly represses transcription by interacting with complexes at the promoter, is bound also to the Sin3 repressor complex, which contains histone deacetylase activities (see **Figure 23.15**). This observation provides a direct connection between two types of repressive modifications: methylation of DNA and acetylation of histones.

The absence of methyl groups is associated with gene expression. However, there are some difficulties in supposing that the state of methylation provides a general means for controlling gene expression. In the case of *D. melanogaster* (and other Dipteran insects), there is very little methylation of DNA (although there is gene potentially coding a methyltransferase), and in the nematode *C. elegans* there is no methylation of DNA. The other differences between inactive and active chromatin appear to be the same as in species that display methylation. So in these organisms, any role that methylation has in vertebrates is replaced by some other mechanism.

We have described three changes that occur in active genes:

- A hypersensitive site (s) is established near the promoter.
- The nucleosomes of a domain including the transcribed region become more sensitive to DNAase I.
- The DNA of the same region is undermethylated.

All of these changes are necessary for transcription.

Last updated on 12-16-2002

Reviews

2424. Bird, A. (2002). *DNA methylation patterns and epigenetic memory*. Genes Dev. 16, 6-21.

References

710. Bird, A. et al. (1985). *A fraction of the mouse genome that is derived from islands of nonmethylated, Cp-G-rich DNA*. Cell 40, 91-99.

711. Boyes, J. and Bird, A. (1991). *DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein*. Cell 64, 1123-1134.

1443. Antequera, F. and Bird, A. (1993). *Number of CpG islands and genes in human and mouse.* Proc. Natl. Acad. Sci. USA 90, 11995-11999.

PROMOTERS AND ENHANCERS

5.21.20 Insulators block the actions of enhancers and heterochromatin

Key Terms

An **insulator** is a sequence that prevents an activating or inactivating effect passing from one side to the other.

Key Concepts

- Insulators are able to block passage of any activating or inactivating effects from enhancers, silencers, or LCRs.
- Insulators may provide barriers against the spread of heterochromatin.

Elements that prevent the passage of activating or inactivating effects are called **insulators** (for review see 2290). They have either or both of two key properties:

- When an insulator is placed between an enhancer and a promoter, *it prevents the enhancer from activating the promoter*. The blocking effect is shown in **Figure 21.29**. This may explain how the action of an enhancer is limited to a particular promoter.
- When an insulator is placed between an active gene and heterochromatin, *it provides a barrier that protects the gene against the inactivating effect that spreads from the heterochromatin*. (Heterochromatin is a region of chromatin that is inactive as the result of its higher order structure; see *Molecular Biology 5.23.13 Heterochromatin propagates from a nucleation event*.) The barrier effect is shown in **Figure 21.30**.

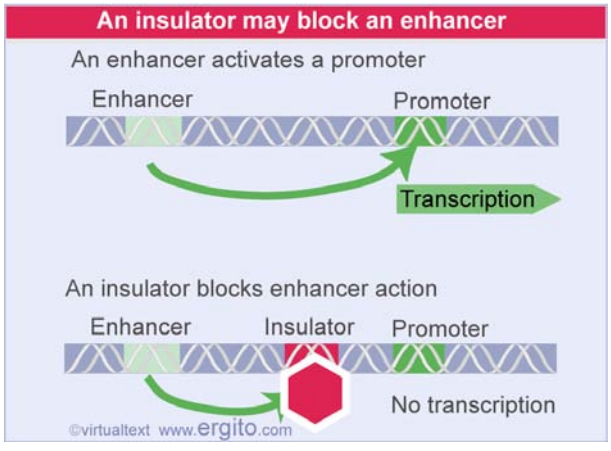

Figure 21.29 An enhancer activates a promoter in its vicinity, but may be blocked from doing so by an insulator located between them.

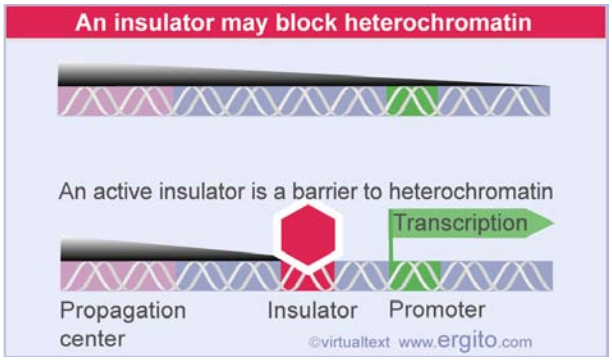

Figure 21.30 Heterochromatin may spread from a center and then blocks any promoters that it covers. An insulator may be a barrier to propagation of heterochromatin that allows the promoter to remain active.

Some insulators possess both these properties, but others have only one, or the blocking and barrier functions can be separated. Although both actions are likely to be mediated by changing chromatin structure, they may involve different effects (for review see 2376). In either case, however, the insulator defines a limit for long-range effects.

What is the purpose of an insulator? A major function may be to counteract the indiscriminate actions of enhancers on promoters. Most enhancers will work with any promoter in the vicinity. An insulator can restrict an enhancer by blocking the effects from passing beyond a certain point, so that it can act only on a specific promoter. Similarly, when a gene is located near heterochromatin, an insulator can prevent it from being inadvertently inactivated by the spread of the heterochromatin. Insulators therefore function as elements for increasing the precision of gene regulation.

Last updated on 3-19-2002

Reviews

2290. Gerasimova, T. I. and Corces, V. G. (2001). *Chromatin insulators and boundaries: effects on transcription and nuclear organization.* Annu. Rev. Genet. 35, 193-208.

2376. West, A. G., Gaszner, M., and Felsenfeld, G. (2002). *Insulators: many functions, many mechanisms.* Genes Dev. 16, 271-288.

PROMOTERS AND ENHANCERS

5.21.21 Insulators can define a domain

Key Concepts

- Insulators are specialized chromatin structures that have hypersensitive sites. Two insulators can protect the region between them from all external effects.

Insulators were discovered during the analysis of the region of the *D. melanogaster* genome summarized in **Figure 21.31**. Two genes for the protein Hsp70 lie within an 18 kb region that constitutes band 87A7. Special structures, called *scs* and *scs ′* (specialized chromatin structures, are found at the ends of the band. Each consists of a region that is highly resistant to degradation by DNAase I, flanked on either side by hypersensitive sites, spaced at about 100 bp. The cleavage pattern at these sites is altered when the genes are turned on by heat shock.

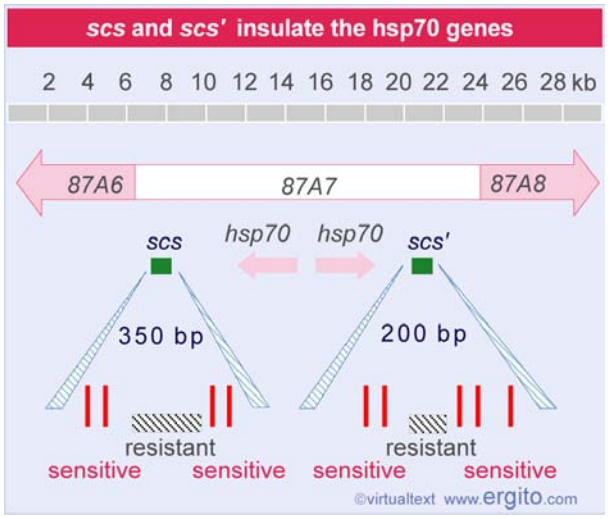

Figure 21.31 Specialized chromatin structures that include hypersensitive sites mark the ends of a domain in the *D. melanogaster* genome and insulate genes between them from the effects of surrounding sequences.

The *scs* elements insulate the *hsp70* genes from the effects of surrounding regions. If we take *scs* units and place them on either side of a *white* gene, the gene can function anywhere it is placed in the genome, even in sites where it would normally be repressed by context, for example, in heterochromatic regions (703).

The *scs* and *scs ′* units do not seem to play either positive or negative roles in controlling gene expression, but just restrict effects from passing from one region to the next. If adjacent regions have repressive effects, however, the *scs* elements might be needed to block the spread of such effects, and therefore could be essential for gene expression. In this case, deletion of such elements could eliminate the expression of the adjacent gene(s).

scs and *scs′* have different structures and each appears to have a different basis for its insulator activity. The key sequence in the *scs* element is a stretch of 24 bp that binds the product of the *zw5* gene (2352). The insulator property of *scs′* resides in a series of CGATA repeats (2350). The repeats bind a group of related proteins called BEAF-32. The protein shows discrete localization within the nucleus, but the most remarkable data derive from its localization on polytene chromosomes. **Figure 21.32** shows that an anti-BEAF-32 antibody stains ~50% of the interbands of the polytene chromosomes (2351). This suggests that there are many insulators in the genome, and that BEAF-32 is a common part of the insulating apparatus. It would imply that the band is a functional unit, and that interbands often have insulators that block the propagation of activating or inactivating effects.

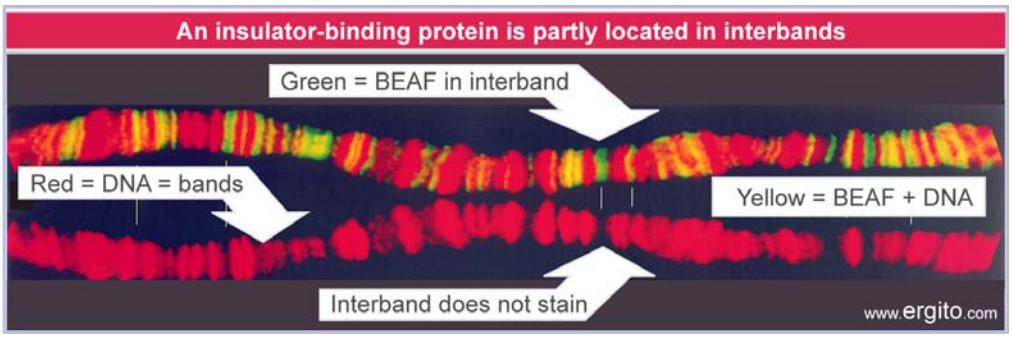

Figure 21.32 A protein that binds to the insulator *scs′* is localized at interbands in *Drosophila* polytene chromosomes. Red staining identifies the DNA (the bands) on both the upper and lower samples; green staining identifies BEAF32 (often at interbands) on the upper sample. Yellow shows coincidence of the two labels (meaning that BEAF32 is in a band). Photograph kindly provided by Uli Laemmli.

Another example of an insulator that defines a domain is found in the chick β-globin LCR (the group of hypersensitive sites that controls expression of all β-globin genes; see *Molecular Biology 5.20.17 An LCR may control a domain*). The leftmost hypersensitive site of the chick β-globin LCR (HS4) is an insulator that marks the 5′ end of the functional domain (705). This restricts the LCR to acting only on the globin genes in the domain.

A gene that is surrounded by insulators is usually protected against the propagation of inactivating effects from the surrounding regions. The test is to inserted DNA into a genome at random locations by transfection. The expression of a gene in the inserted sequence is often erratic; in some instances it is properly expressed, but in others it is extinguished (see *Molecular Biology 4.18.18 Genes can be injected into animal eggs*). However, when insulators that have a barrier function are placed on either side of the gene in the inserted DNA, its expression typically is uniform in every case (3216).

References

703. Kellum, R. and Schedl, P. (1991). *A position-effect assay for boundaries of higher order chromosomal domains*. Cell 64, 941-950.

705. Chung, J. H., Whiteley, M., and Felsenfeld, G. (1993). *A 5′ element of the chicken β -globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila*. Cell 74, 505-514.

2350. Cuvier, O., Hart, C. M., and Laemmli, U. K. (1998). *Identification of a class of chromatin boundary elements.* Mol. Cell Biol. 18, 7478-7486.

2351. Zhao, K., Hart, C. M., and Laemmli, U. K. (1995). *Visualization of chromosomal domains with boundary element-associated factor BEAF-32.* Cell 81, 879-889.

2352. Gaszner, M., Vazquez, J., and Schedl, P. (1999). *The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction.* Genes Dev. 13, 2098-2107.

3216. Pikaart, M. J., Recillas-Targa, F., and Felsenfeld, G. (1998). *Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators.* Genes Dev. 12, 2852-2862.

PROMOTERS AND ENHANCERS

5.21.22 Insulators may act in one direction

Key Concepts

- Some insulators have directionality, and may stop passage of effects in one direction but not the other.

Insulators may have directional properties. Insertions of the transposon *gypsy* into the *yellow* (*y*) locus of *D. melanogaster* cause loss of gene function in some tissues, but not in others. The reason is that the *y* locus is regulated by four enhancers, as shown in **Figure 21.33**. Wherever *gypsy* is inserted, it blocks expression of all enhancers that it separates from the promoter, but not those that lie on the other side. The sequence responsible for this effect is an insulator that lies at one end of the transposon. The insulator works irrespective of its orientation of insertion.

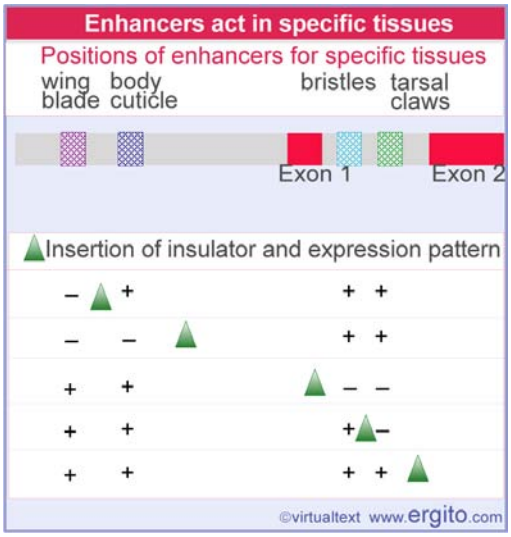

Figure 21.33 The insulator of the gypsy transposon blocks the action of an enhancer when it is placed between the enhancer and the promoter.

Some of the enhancers are upstream of the promoter and others are downstream, so the effect cannot depend on position with regard to the promoter, nor can it require transcription to occur through the insulator. This is difficult to explain in terms of looping models for enhancer-promoter interaction, which essentially predict the irrelevance of the intervening DNA. The obvious model to invoke is a tracking mechanism, in which some component must move unidirectionally from the enhancer to the promoter, but this is difficult to reconcile with previous characterizations of the independence of enhancers from such effects.

Proteins that act upon the insulator have been identified through the existence of two other loci that affect insulator function in a *trans*-acting manner. Mutations in *su(Hw)*

abolish insulation: *y* is expressed in all tissues in spite of the presence of the insulator. This suggests that *su(Hw)* codes for a protein that recognizes the insulator and is necessary for its action. Su(Hw) has a zinc finger DNA-motif; mapping to polytene chromosomes shows that it is bound at a large number of sites. The insulator contains 12 copies of a 26 bp sequence that is bound by Su(Hw) (706; 709). Manipulations show that the strength of the insulator is determined by the number of copies of the binding sequence.

The second locus is *mod(mdg4)*, in which mutations have the opposite effect. This is observed by the loss of directionality. These mutations increase the effectiveness of the insulator by extending its effects so that it blocks utilization of enhancers on both sides. *su(Hw)* is epistatic to *mod(mdg4)*; this means that in a double mutant we see only the effect of *su(Hw)*. This implies that mod(mdg4) acts through su(Hw). The basic role of the wild-type protein from the *mod(mdg4)* locus is therefore to impose directionality on the ability of su(Hw) to insulate promoters from the boundary.

Binding of su(Hw) to DNA, followed by binding of mod(mdg4) to su(Hw), therefore creates a unidirectional block to activation of a promoter. This suggests that the insulator bound by su(Hw) can spread inactivity in both directions, but mod(mdg4) stops the effect from spreading in one direction. Perhaps there is some intrinsic directionality to chromatin, which results ultimately in the incorporation of su(Hw), mod(mdg4), or some other component in one orientation, presumably by virtue of an interaction with some component of chromatin that is itself preferentially oriented. Any such directionality would need to reverse at the promoter.

It is likely that insulators act by making changes in chromatin structure. One model is prompted by the observation that Su(Hw) and mod(mdg4) binding sites are present at >500 locations in the Drosophila genome. But visualization of the sites where the proteins are bound in the nucleus shows that they are colocalized at ~25 discrete sites around the nuclear periphery (2377). This suggests the model of **Figure 21.34** in which Su(Hw) proteins bound at different sites on DNA are brought together by binding to mod(mdg4). The Su(Hw)/mod(mdg4) complex is localized at the nuclear periphery. The DNA bound to it is organized into loops. An average complex might have ~20 such loops. Enhancer-promoter actions can occur only within a loop, and cannot propagate between them.

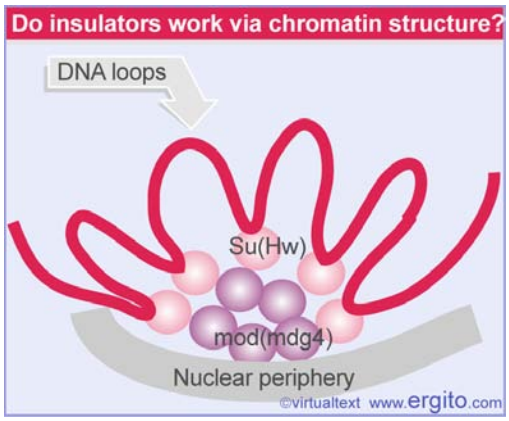

Figure 21.34 Su(Hw)/mod(mdg4) complexes are found at the nuclear periphery. They may organize DNA into loops that limit enhancer-promoter interactions.

Last updated on 3-15-2002

References

706. Roseman, R. R., Pirrotta, V., and Geyer, P. K. (1993). *The su(Hw) protein insulates expression of the D. melanogaster white gene from chromosomal position-effects.* EMBO J. 12, 435-442.

709. Harrison, D. A., Gdula, D. A., Cyne, R. S., and Corces, V. G. (1993). *A leucine zipper domain of the suppressor of hairy-wing protein mediates its repressive effect on enhancer function Genes Dev*. Genes Dev. 7, 1966-1978.

2377. Gerasimova, T. I., Byrd, K., and Corces, V. G. (2000). *A chromatin insulator determines the nuclear localization of DNA.* Mol. Cell 6, 1025-1035.

PROMOTERS AND ENHANCERS

5.21.23 Insulators can vary in strength

Key Terms

An **anti-insulator** is a sequence that allows an enhancer to overcome the effect of an insulator.

Key Concepts

- Insulators can differ in how effectively they block passage of an activating signal.

Sometimes elements with different *cis*-acting properties are combined to generate regions with complex regulatory effects. The *Fab-7* region is defined by deletions in the *bithorax* locus of *Drosophila*. This locus contains a series of *cis*-acting regulatory elements that control the activities of three transcription units (see **Figure 31.36**). The relevant part of the locus is drawn in **Figure 21.35**. The regulatory elements *iab-6* and *iab-7* control expression of the adjacent gene *Abd-B* in successive regions of the embryo (segments A6 and A7). A deletion of *Fab-7* causes A6 to develop like A7, instead of in the usual way. This is a dominant effect, which suggests that *iab-7* has taken over control from *iab-6*. We can interpret this in molecular terms by supposing that *Fab-7* provides a boundary that prevents *iab-7* from acting when *iab-6* is usually active (707; 708).

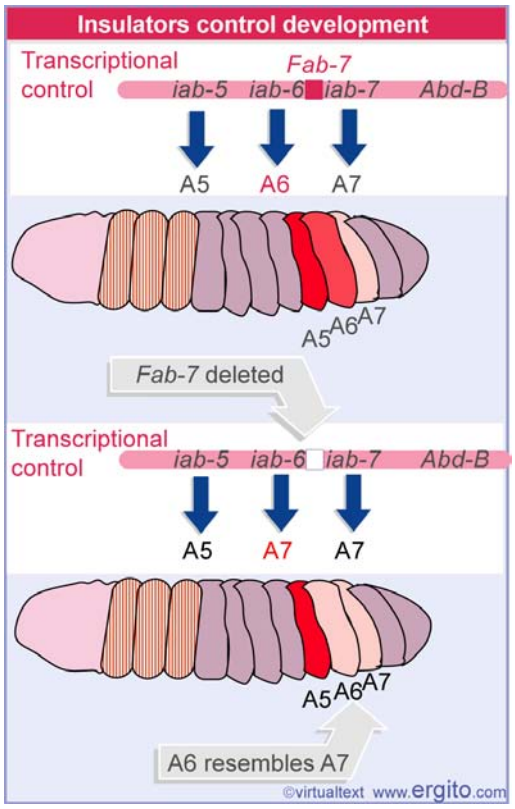

Figure 21.35 *Fab-7* is a boundary element that is necessary for the independence of regulatory elements *iab-6* and *iab-7*.

Like other boundary elements, *Fab-7* contains a distinctive chromatin structure that is marked by a series of hypersensitive sites. The region can be divided into two types of elements by smaller deletions and by testing fragments for their ability to provide a boundary. A sequence of ~3.3 kb behaves as an insulator when it is placed in other constructs. A sequence of ~0.8 kb behaves as a repressor that acts on *iab-7*. The presence of these two elements explains the complicated genetic behavior of *Fab-7* (which we have not described in detail).

An insight into the action of the boundary element is provided by the effects of substituting other insulators for *Fab-7*. The effect of *Fab-7* is simply to prevent interaction between *iab-6* and *iab-7*. But when *Fab-7* is replaced by a different insulator [in fact a binding site for the protein Su(Hw)], a stronger effect is seen: *iab-5* takes over from *iab-7*. And when an *scs* element is used, the effect extends to *iab-4*. This suggests a scheme in which stronger elements can block the actions of regulatory sequences that lie farther away.

This conclusion introduces a difficulty for explaining the action of boundary elements. They cannot be functioning in this instance simply by preventing the transmission of effects past the boundary. This argues against models based on simple tracking or inhibiting the linear propagation of structural effects. It suggests that there may be some sort of competitive effect, in which the strength of the element determines how far its effect can stretch.

The situation is further complicated by the existence of **anti-insulator** elements, which allow an enhancer to overcome the blocking effects of an insulator (940). This again suggests that these effects are mediated by some sort of control over local chromatin structure.

Last updated on 2-15-2002

References

707. Hagstrom, K., Muller, M., and Schedl, P. (1996). *Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex*. Genes Dev. 10, 3202-3215.

708. Mihaly, J. et al. (1997). *In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element*. Development 124, 1809-1820.

940. Zhou, J. and Levine, M. (1999). *A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the Drosophila embryo.* Cell 99, 567-575.

PROMOTERS AND ENHANCERS

5.21.24 What constitutes a regulatory domain?

Key Terms

An **insulator** is a sequence that prevents an activating or inactivating effect passing from one side to the other.

A **matrix attachment site (MAR)** is a region of DNA that attaches to the nuclear matrix. It is also known as a scaffold attachment site (SAR).

The **locus control region (LCR)** that is required for the expression of several genes in a domain.

Key Concepts

- A domain has an insulator, an LCR, a matrix attachment site, and transcription unit(s).

If we now put together the various types of structures that have been found in different systems, we can think about the possible nature of a chromosomal domain. The basic feature of a regulatory domain is that regulatory elements can act only on transcription units within the same domain. A domain might contain more than one transcription unit and/or enhancer.

Figure 21.36 summarizes the structures that might be involved in defining a domain.

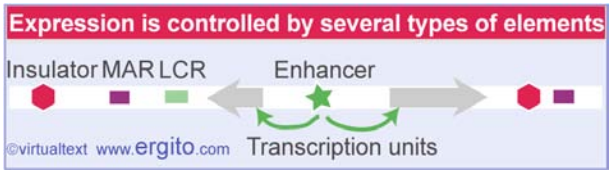

Figure 21.36 Domains may possess three types of sites: insulators to prevent effects from spreading between domains; MARs to attach the domain to the nuclear matrix; and LCRs that are required for initiation of transcription. An enhancer may act on more than one promoter within the domain.

An **insulator** stops activating or repressing effects from passing. In its simplest form, an insulator blocks either type of effect from passing across it, but there can be more complex relationships in which the insulator blocks only one type of effect and/or acts directionally. We assume that insulators act by affecting higher order chromatin structure, but we do not know the details and varieties of such effects.

A **matrix attachment site** (**MAR**) may be responsible for attaching chromatin to a site on the nuclear periphery (see *Molecular Biology 5.19.6 Specific sequences attach DNA to an interphase matrix*). These are likely to be responsible for creating physical domains of DNA that take the form of loops extending out from the attachment sites. This looks very like one model for insulator action. In fact, some

MAR elements behave as insulators in assays *in vitro*, but it seems that the their ability to attach DNA to the matrix can be separated from the insulator function, so there is not a simple cause and effect. It would not be surprising if insulator and MAR elements were associated to maintain a relationship between regulatory effects and physical structure (for review see 2376).

An **LCR** functions at a distance and may be required for any and all genes in a domain to be expressed (see *Molecular Biology 5.20.17 An LCR may control a domain*). When a domain has an LCR, its function is essential for all genes in the domain, but LCRs do not seem to be common. Several types of *cis*-acting structures could be required for function. As defined originally, the property of the LCR rests with an enhancer-like hypersensitive site that is needed for the full activity of promoter(s) within the domain.

The organization of domains may help to explain the large size of the genome. A certain amount of space could be required for such a structure to operate, for example, to allow chromatin to become decondensed and to become accessible. Although the exact sequences of much of the unit might be irrelevant, there might be selection for the overall amount of DNA within it, or at least selection might prevent the various transcription units from becoming too closely spaced.

Last updated on 3-19-2002

Reviews

2376. West, A. G., Gaszner, M., and Felsenfeld, G. (2002). *Insulators: many functions, many mechanisms.* Genes Dev. 16, 271-288.

PROMOTERS AND ENHANCERS

5.21.25 Summary

Of the three eukaryotic RNA polymerases, RNA polymerase I transcribes rDNA and accounts for the majority of activity, RNA polymerase II transcribes structural genes for mRNA and has the greatest diversity of products, and RNA polymerase III transcribes small RNAs. The enzymes have similar structures, with two large subunits and many smaller subunits; there are some common subunits among the enzymes.

None of the three RNA polymerases recognize their promoters directly. A unifying principle is that transcription factors have primary responsibility for recognizing the characteristic sequence elements of any particular promoter, and they serve in turn to bind the RNA polymerase and to position it correctly at the startpoint. At each type of promoter, the initiation complex is assembled by a series of reactions in which individual factors join (or leave) the complex. The factor TBP is required for initiation by all three RNA polymerases. In each case it provides one subunit of a transcription factor that binds in the vicinity of the startpoint.

A promoter consists of a number of short sequence elements in the region upstream of the startpoint. Each element is bound by a transcription factor. The basal apparatus, which consists of the TF_{II} factors, assembles at the startpoint and enables RNA polymerase to bind. The TATA box (if there is one) near the startpoint, and the initiator region immediately at the startpoint, are responsible for selection of the exact startpoint at promoters for RNA polymerase II. TBP binds directly to the TATA box when there is one; in TATA-less promoters it is located near the startpoint by binding to the DPE downstream. After binding of $TF_{II}D$, the other general transcription factors for RNA polymerase II assemble the basal transcription apparatus at the promoter. Other elements in the promoter, located upstream of the TATA box, bind activators that interact with the basal apparatus. The activators and basal factors are released when RNA polymerase begins elongation.

The CTD of RNA polymerase II is phosphorylated during the initiation reaction. $TF_{II}D$ and SRB proteins both may interact with the CTD. It may also provide a point of contact for proteins that modify the RNA transcript, including the 5′ capping enzyme, splicing factors, and the 3′ processing complex.

Promoters may be stimulated by enhancers, sequences that can act at great distances and in either orientation on either side of a gene. Enhancers also consist of sets of elements, although they are more compactly organized. Some elements are found in both promoters and enhancers. Enhancers probably function by assembling a protein complex that interacts with the proteins bound at the promoter, requiring that DNA between is "looped out."

An insulator blocks the transmission of activating or inactivating effects in chromatin. An insulator that is located between an enhancer and a promoter prevents the enhancer from activating the promoter. Two insulators define the region between them as a regulatory domain; regulatory interactions within the domain are limited to it, and the domain is insulated from outside effects. Most insulators block regulatory

effects from passing in either direction, but some are directional. Insulators usually can block both activating effects (enhancer-promoter interactions) and inactivating effects (mediated by spread of heterochromatin), but some are limited to one or the other. Insulators are thought to act via changing higher order chromatin structure, but the details are not certain.

CpG islands contain concentrations of CpG doublets and often surround the promoters of constitutively expressed genes, although they are also found at the promoters of regulated genes. The island including a promoter must be unmethylated for that promoter to be able to initiate transcription. A specific protein binds to the methylated CpG doublets and prevents initiation of transcription.