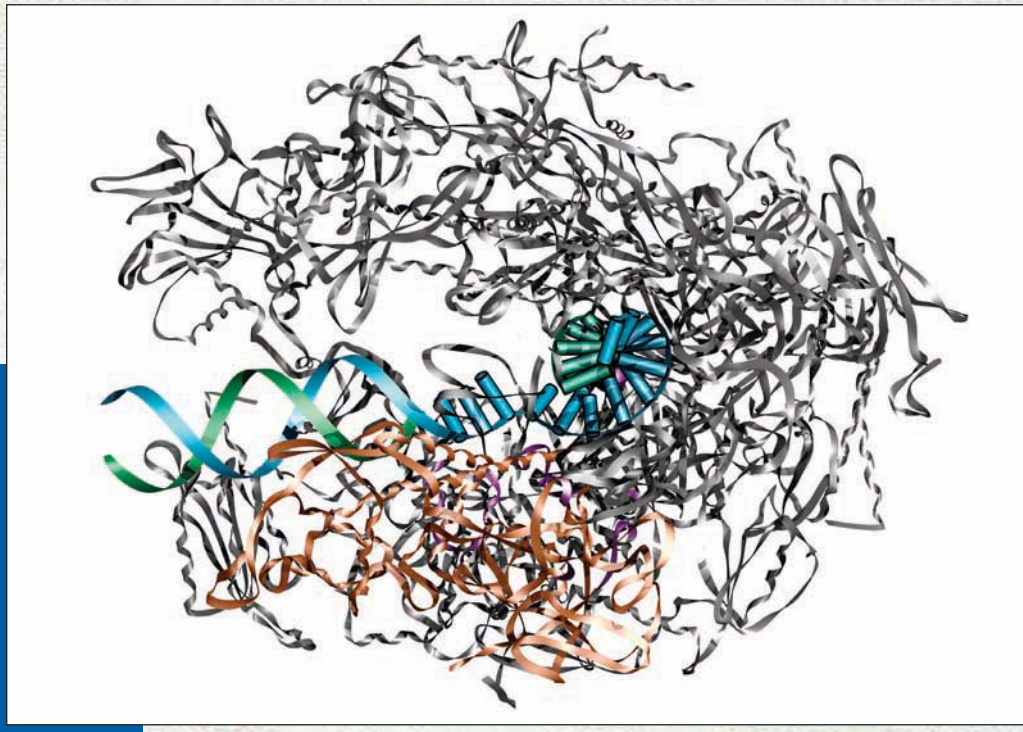
CHAPTER 10

Eukaryotic RNA Polymerases and Their Promoters



Computer-generated model of yeast Pol II Δ4/7 protein with RNA–DNA hybrid in the active site. *© David A. Bushnell, Kenneth D. Westover, and Roger D. Kornberg.*

In Chapter 6 we learned that bacteria have only one RNA polymerase, which makes all three of the familiar RNA types: mRNA, rRNA, and tRNA. True, the polymerase can switch σ-factors to meet the demands of a changing environment, but the core enzyme remains essentially the same. Quite a different situation prevails in the eukaryotes. In this chapter we will see that three distinct RNA polymerases occur in the nuclei of eukaryotic cells. Each of these is responsible for transcribing a separate set of genes, and each recognizes a different kind of promoter.

10.1 Multiple Forms of Eukaryotic RNA Polymerase

Several early studies suggested that at least two RNA polymerases operate in eukaryotic nuclei: one to transcribe the major ribosomal RNA genes (those coding for the 28S, 18S, and 5.8S rRNAs in vertebrates), and one or more to transcribe the rest of the nuclear genes.

To begin with, the ribosomal genes are different in several ways from other nuclear genes: (1) They have a different base composition from that of other nuclear genes. For example, rat rRNA genes have a GC content of 60%, but the rest of the DNA has a GC content of only 40%. (2) They are unusually repetitive; depending on the organism, each cell contains from several hundred to over 20,000 copies of the rRNA gene. (3) They are found in a different compartment—the nucleolus—than the rest of the nuclear genes. These and other considerations suggested that at least two RNA polymerases were operating in eukaryotic nuclei. One of these synthesized rRNA in the nucleolus, and the other synthesized other RNA in the nucleoplasm (the part of the nucleus outside the nucleolus).

Separation of the Three Nuclear Polymerases

Robert Roeder and William Rutter showed in 1969 that eukaryotes have not two, but three different RNA polymerases. Furthermore, these three enzymes have distinct roles in the cell. These workers separated the three enzymes by DEAE-Sephadex ion-exchange chromatography (Chapter 5).

They named the three peaks of polymerase activity in order of their emergence from the ion-exchange column: **RNA polymerase I, RNA polymerase II,** and **RNA polymerase III** (Figure 10.1). The three enzymes have different properties besides their different behaviors on DEAE-Sephadex chromatography. For example, they have different responses to ionic strength and divalent metals. More importantly, they have distinct roles in transcription: Each makes different kinds of RNA.

Roeder and Rutter next looked in purified nucleoli and nucleoplasm to see if these subnuclear compartments were enriched in the appropriate polymerases. Figure 10.2 shows that polymerase I is indeed located primarily in the nucleolus, and polymerases II and III are found in the nucleoplasm. This made it very likely that polymerase I is the rRNA-synthesizing enzyme, and that polymerases II and III make some other kinds of RNA.

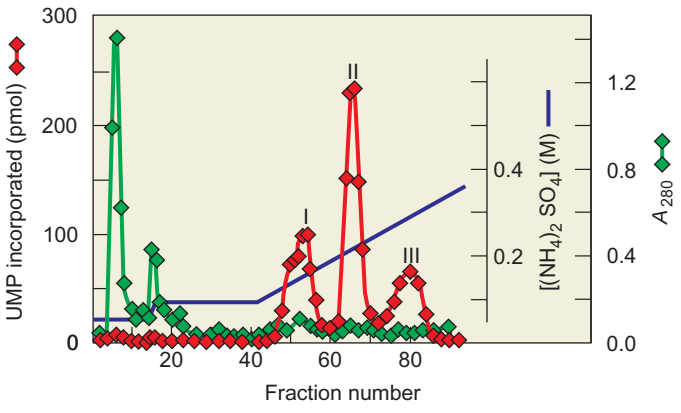


Figure 10.1 Separation of eukaryotic RNA polymerases. Roeder and Rutter subjected extracts from sea urchin embryos to DEAE-Sephadex chromatography. Green, protein measured by A_{280}; red, RNA polymerase activity measured by incorporation of labeled UMP into RNA; blue, ammonium sulfate concentration. (*Source:* Adapted from Roeder, R.G. and W.J. Rutter, Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224:235, 1969.)

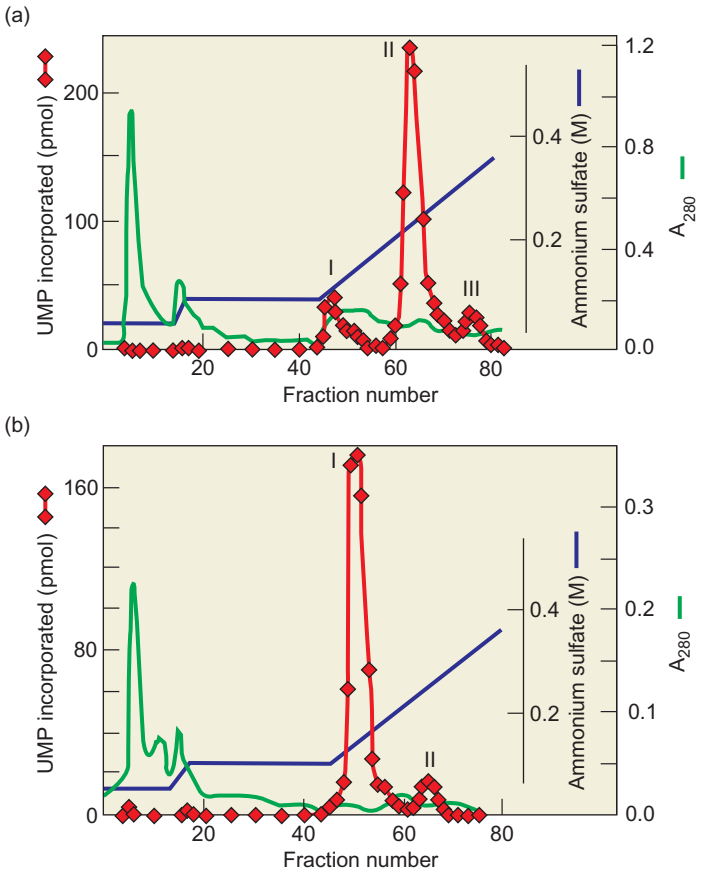


Figure 10.2 Cellular localization of the three rat liver RNA polymerases. Roeder and Rutter subjected the polymerases found in the nucleoplasmic fraction **(a)** or nucleolar fraction **(b)** of rat liver to DEAE-Sephadex chromatography as described in Figure 10.1. Colors have the same meanings as in Figure 10.1. (*Source:* Adapted from Roeder, R.G. and W.J. Rutter, Specific nucleolar and nucleoplasmic RNA polymerases, *Proceedings of the National Academy of Sciences* 65(3):675–82, March 1970.)

SUMMARY Eukaryotic nuclei contain three RNA polymerases that can be separated by ion-exchange chromatography. RNA polymerase I is found in the nucleolus; the other two polymerases (RNA polymerases II and III) are located in the nucleoplasm. The location of RNA polymerase I in the nucleolus suggests that it transcribes the rRNA genes.

The Roles of the Three RNA Polymerases

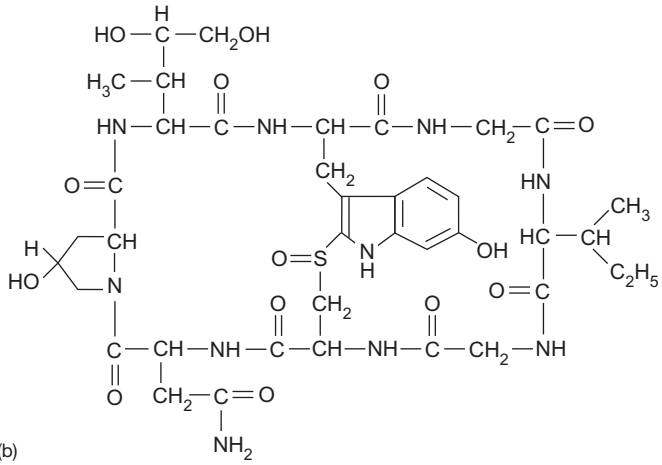
How do we know that the three RNA polymerases have different roles in transcription? The clearest evidence for these roles has come from studies in which the purified polymerases were shown to transcribe certain genes, but not others, in vitro. Such studies have demonstrated that the three RNA polymerases have the following specificities (Table 10.1): Polymerase I makes the large rRNA precursor. In mammals, this precursor has a sedimentation coefficient of 45S and is processed to the 28S, 18S, and 5.8S mature rRNAs. Polymerase II makes an ill-defined class of RNA known as **heterogeneous nuclear RNA** (**hnRNA**) as well as the precursors of **microRNAs** (**miRNAs**) and most **small nuclear RNAs** (**snRNAs**). We will see in Chapter 14 that most of the hnRNAs are precursors of mRNAs and that the snRNAs participate in the maturation of hnRNAs to mRNAs. In Chapter 16, we will learn that microRNAs control the expression of many genes by causing degradation of, or limiting the translation of, their mRNAs. Polymerase III makes precursors to the tRNAs, 5S rRNA, and some other small RNAs.

However, even before cloned genes and eukaryotic in vitro transcription systems were available, we had evidence to support most of these transcription assignments. In this section, we will examine the early evidence that RNA polymerase III transcribes the tRNA and 5S rRNA genes.

Table 10.1 Roles of Eukaryotic RNA Polymerases

RNA Polymerase	Cellular RNAs Synthesized	Mature RNA (Vertebrate)
I	Large rRNA precursor	28S, 18S, and 5.8S rRNAs
II	hnRNAs	mRNAs
	snRNAs	snRNAs
	miRNA precursors	miRNAs
III	5S rRNA precursor	5S rRNA
	tRNA precursors	tRNAs
	U6 snRNA (precursor?)	U6 snRNA
	7SL RNA (precursor?)	7SL RNA
	7SK RNA (precursor?)	7SK RNA

(a)



(b)

Figure 10.3 Alpha-amanitin. (a) *Amanita phalloides* ("the death cap"), one of the deadly poisonous mushrooms that produce α-amanitin. **(b)** Structure of α-amanitin. (*Source: (a)* Arora, D. *Mushrooms Demystified* 2e, 1986, Plate 50 (Ten Speed Press).)

This work, by Roeder and colleagues in 1974, depended on a toxin called **α-amanitin.** This highly toxic substance is found in several poisonous mushrooms of the genus *Amanita* (Figure 10.3a), including *A. phalloides,* "the death cap," and *A. bisporigera,* which is called "the angel of death" because it is pure white and deadly poisonous. Both species have proven fatal to many inexperienced mushroom hunters. Alpha-amanitin was found to have different effects on the three polymerases. At very low concentrations, it inhibits polymerase II completely while having no effect at all on polymerases I and III. At 1000-fold higher concentrations, the toxin also inhibits polymerase III from most eukaryotes (Figure 10.4).

The plan of the experiment was to incubate mouse cell nuclei in the presence of increasing concentrations of α-amanitin, then to electrophorese the transcripts to observe the effect of the toxin on the synthesis of small RNAs. Figure 10.5 reveals that high concentrations of α-amanitin inhibited the synthesis of both 5S rRNA and 4S tRNA

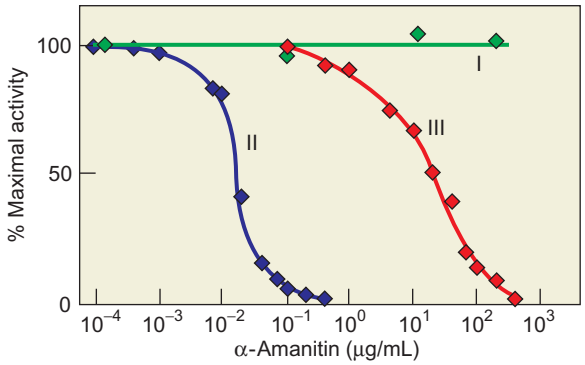


Figure 10.4 Sensitivity of purified RNA polymerases to α-amanitin. Weinmann and Roeder assayed RNA polymerases I (green), II (blue), and III (red) with increasing concentrations of α-amanitin. Polymerase II was 50% inhibited by about 0.02 μg/mL of the toxin, whereas polymerase III reached 50% inhibition only at about 20 μg/mL of toxin. Polymerase I retained full activity even at an α-amanitin concentration of 200 μg/mL. (*Source:* Adapted From R. Weinmann and R.G. Roeder, Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes, *Proceedings of the National Academy of Sciences USA* 71(5):1790–4, May 1974.)

precursor. Moreover, this pattern of inhibition of 5S rRNA and tRNA precursor synthesis matched the pattern of inhibition of RNA polymerase III: They both were about half-inhibited at 10 μg/mL of α-amanitin. Therefore, these data support the hypothesis that RNA polymerase III makes these two kinds of RNA. (Actually, polymerase III synthesizes the 5S rRNA as a slightly larger precursor, but this experiment did not distinguish the precursor from the mature 5S rRNA.) Polymerase III also makes a variety of other small cellular and viral RNAs. These include **U6 snRNA,** a small RNA that participates in RNA splicing (Chapter 14); **7SL RNA,** a small RNA involved in signal peptide recognition in the synthesis of secreted proteins; 7SK RNA, a small nuclear RNA that binds and inhibits the class II transcription elongation factor P-TEFb, the adenovirus VA (virus-associated) RNAs; and the Epstein–Barr virus EBER2 RNA.

Similar experiments were performed to identify the genes transcribed by RNA polymerases I and II. But these studies were not as easy to interpret and they have been confirmed by much more definitive in vitro studies.

The sequencing of the first plant genome (*Arabidopsis thaliana*, or thale cress) in 2000 led to the discovery of two

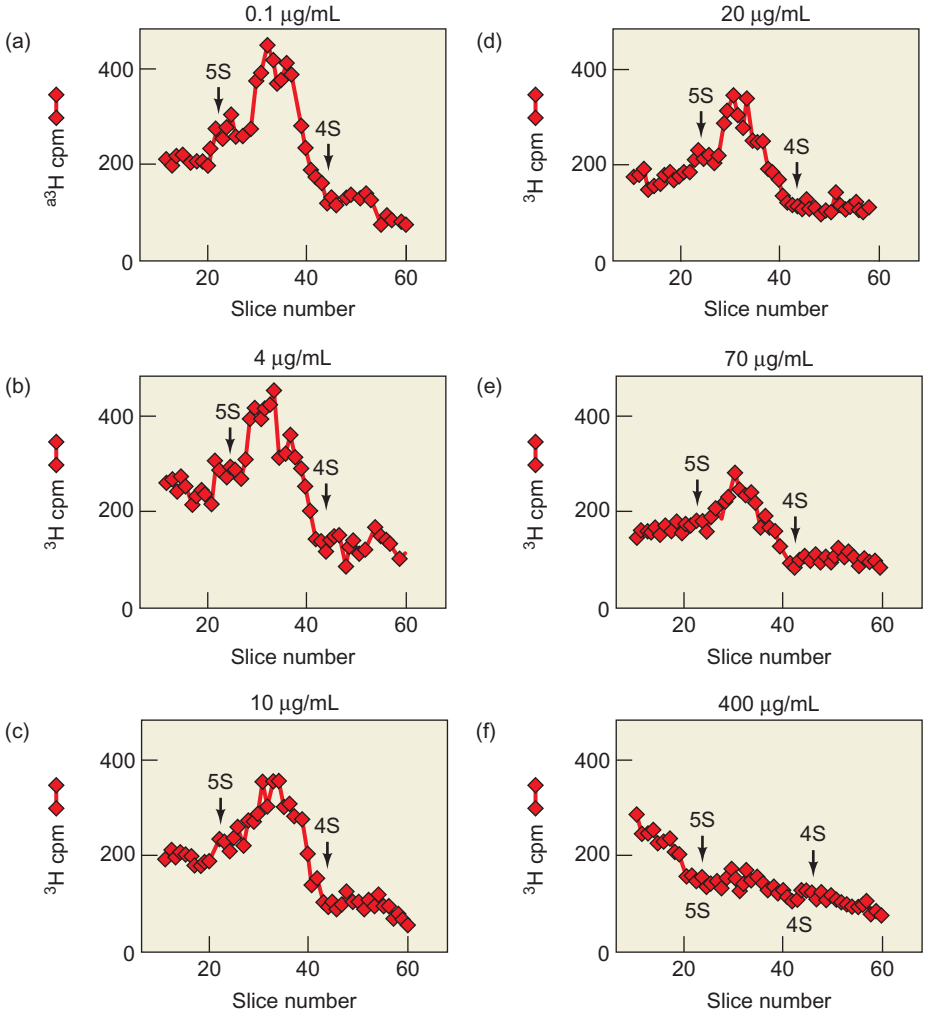


Figure 10.5 Effect of α-amanitin on small RNA synthesis. Weinmann and Roeder synthesized labeled RNA in isolated nuclei in the presence of increasing amounts of α-amanitin (concentration given at the top of each panel). The small labeled RNAs leaked out of the nuclei and were found in the supernatant after centrifugation. The researchers then subjected these RNAs to PAGE, sliced the gel, and determined the radioactivity in each slice (red). They also ran markers (5S rRNA and 4S tRNA) in adjacent lanes of the same gel. The inhibition of 5S rRNA and 4S tRNA precursor synthesis by α-amanitin closely parallels the effect of the toxin on polymerase III, determined in Figure 10.4. (*Source:* Adapted from R. Weinmann and R.G. Roeder, Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes, *Proceedings of the National Academy of Sciences USA* 71(5):1790–4, May 1974.)

additional RNA polymerases in flowering plants: **RNA polymerase IV** and **RNA polymerase V.** These enzymes produce noncoding RNAs that are involved in a mechanism that silences genes. (Similar transcriptional tasks are performed by polymerase II in other eukaryotes, and indeed the largest subunits of both polymerases IV and V are evolutionarily related to the largest subunit of polymerase II.) We will discuss such gene silencing mechanisms in more detail in Chapter 16.

SUMMARY The three nuclear RNA polymerases have different roles in transcription. Polymerase I makes the large precursor to the rRNAs (5.8S, 18S, and 28S rRNAs in vertebrates). Polymerase II makes hnRNAs, which are precursors to mRNAs, miRNA precursors, and most of the snRNAs. Polymerase III makes the precursors to 5S rRNA, the tRNAs, and several other small cellular and viral RNAs.

RNA Polymerase Subunit Structures

The first subunit structures for a eukaryotic RNA polymerase (polymerase II) were reported independently by Pierre Chambon and Rutter and their colleagues in 1971, but they were incomplete. We should note in passing that Chambon named his three polymerases A, B, and C, instead of I, II, and III, respectively. However, the I, II, III nomenclature of Roeder and Rutter has become the standard. We now have very good structural information on all three polymerases from a variety of eukaryotes. The structures of all three polymerases are quite complex, with 14, 12, and 17 subunits in polymerases I, II, and III, respectively. Polymerase II is by far the best studied, and we will focus the rest of our discussion on the structure and function of that enzyme.

Polymerase II Structure For enzymes as complex as the eukaryotic RNA polymerases it is difficult to tell which polypeptides that copurify with the polymerase activity are really subunits of the enzymes and which are merely contaminants that bind tightly to the enzymes. One way of dealing with this problem would be to separate the putative subunits of a polymerase and then see which polypeptides are really required to reconstitute polymerase activity. Although this strategy worked beautifully for the prokaryotic polymerases, no one has yet been able to reconstitute a eukaryotic nuclear polymerase from its separate subunits. Thus, one must try a different tack.

Another way of approaching this problem is to find the genes for all the putative subunits of a polymerase, mutate them, and determine which are required for activity. This has been accomplished for one enzyme: polymerase II of baker's yeast, *Saccharomyces cerevisiae*. Several investigators used traditional methods to purify yeast polymerase II to homogeneity and identified 10 putative subunits. Later, some of the same scientists discovered two other subunits that had been hidden in the earlier analyses, so the current concept of the structure of yeast polymerase II includes 12 subunits. The genes for all 12 subunits have been sequenced, which tells us the amino acid sequences of their products. The genes have also been systematically mutated, and the effects of these mutations on polymerase II activity have been observed.

Table 10.2 lists the 12 subunits of human and yeast polymerase II, along with their molecular masses and some of

Table 10.2 Human and Yeast RNA Polymerase II Subunits

Subunit	Yeast Gene	Yeast Protein (kD)	Features
hRPB1	*RPB1*	192	Contains CTD; binds DNA; involved in start site selection; β′ ortholog
hRPB2	*RPB2*	139	Contains active site; involved in start site selection, elongation rate; β ortholog
hRPB3	*RPB3*	35	May function with Rpb11 as ortholog of the α dimer of prokaryotic RNA polymerase
hRPB4	*RPB4*	25	Subcomplex with Rpb7; involved in stress response
hRPB5	*RPB5*	25	Shared with Pol I, II, III; target for transcriptional activators
hRPB6	*RPB6*	18	Shared with Pol I, II, III; functions in assembly and stability
hRPB7	*RPB7*	19	Forms subcomplex with Rpb4 that preferentially binds during stationary phase
hRPB8	*RPB8*	17	Shared with Pol I, II, III; has oligonucleotide/oligosaccharide-binding domain
hRPB9	*RPB9*	14	Contains zinc ribbon motif that may be involved in elongation: functions in start site selection
hRPB10	*RPB10*	8	Shared with Pol I, II, III
hRPB11	*RPB11*	14	May function with Rpb3 as ortholog of the α dimer of prokaryotic RNA polymerase
hRPB12	*RPB12*	8	Shared with Pol I, II, III

their characteristics. Each of these polypeptides is encoded in a single gene in the yeast and human genomes. The names of these polymerase subunits, Rpb1, and so on, derive from the names of the genes that encode them (*RPB1,* and so on). Note the echo of the Chambon nomenclature in the name *RPB,* which stands for RNA polymerase B (or II).

How do the structures of polymerases I and III compare with this polymerase II structure? First, all the polymerase structures are complex—even more so than the structures of the bacterial polymerases. Second, all the structures are similar in that each contains two large (greater than 100 kD) subunits, plus a variety of smaller subunits. In this respect, these structures resemble those of the prokaryotic core polymerases, which contain two high-molecular-mass subunits (β and β′) plus three low-molecular-mass subunits (two α's and an ω). In fact, as we will see later in this chapter, an evolutionary relationship is evident between three of the prokaryotic core polymerase subunits and three of the subunits of all of the eukaryotic polymerases. In other words, the three eukaryotic polymerases are related to the prokaryotic polymerase and to one another.

A third message from Table 10.2 is that the three yeast nuclear polymerases have several subunits in common. In fact, five such *common subunits* exist. In the polymerase II structure, these are called Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12. These are identified on the right in Table 10.2.

Richard Young and his coworkers originally identified 10 polypeptides that are authentic polymerase II subunits, or at least tightly bound contaminants. The method they used is called **epitope tagging** (Figure 10.6), in which they attached a small foreign epitope to one of the yeast polymerase II subunits (Rpb3) by engineering its gene. Then they introduced this gene into yeast cells lacking a functional Rpb3 gene, labeled the cellular proteins with either ^{35}S or ^{32}P, and used an antibody directed against the foreign epitope to precipitate the whole enzyme. After immunoprecipitation, they separated the labeled polypeptides of the precipitated protein by SDS-PAGE and detected them by autoradiography. Figure 10.7a presents the results. This single-step purification method yielded essentially pure polymerase II with 10 apparent subunits. We can also see a few minor polypeptides, but they are equally visible in the control in which wild-type enzyme, with no epitope tag, was used. Therefore, they are not polymerase-associated. Figure 10.7b shows a later SDS-PAGE analysis of the same polymerase, performed by Roger Kornberg and colleagues, which distinguished 12 subunits. Rpb11 had coelectrophoresed with Rpb9, and Rpb12 had coelectrophoresed with Rpb10, so both Rpb11 and Rpb12 had been missed in the earlier experiments.

Because Young and colleagues already knew the amino acid compositions of all 10 original subunits, the relative labeling of each polypeptide with ^{35}S-methionine gave them a good estimate of the stoichiometries of subunits, which are listed in Table 10.3. Figure 10.7a also shows us that two

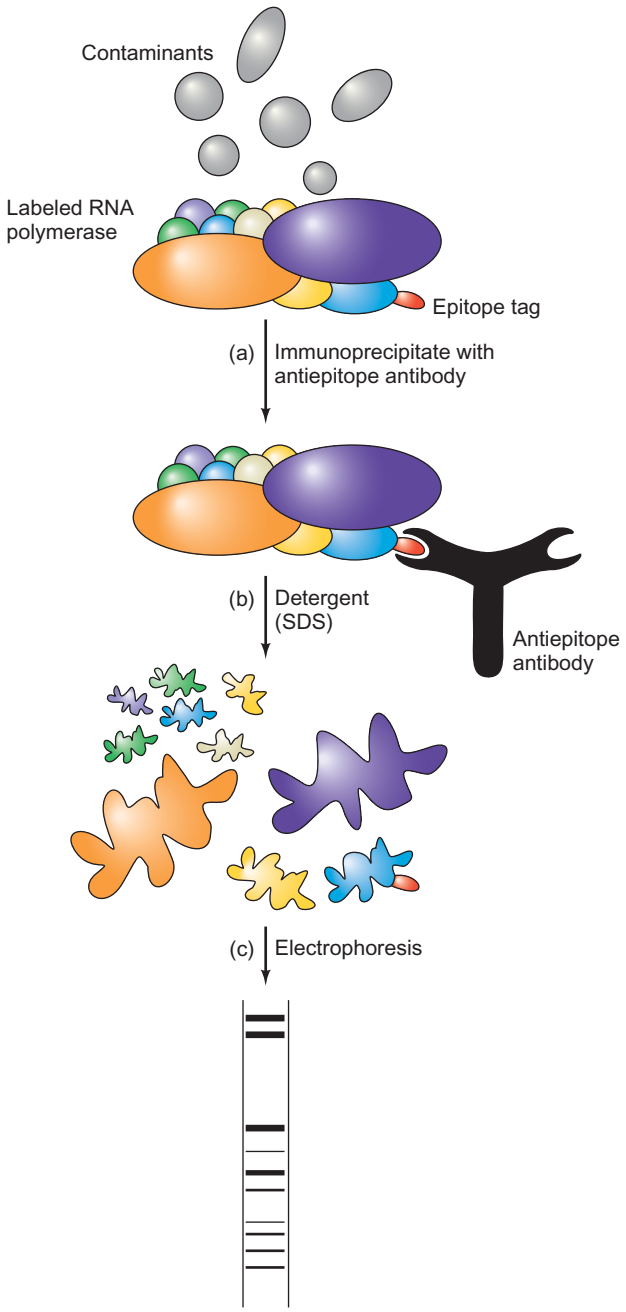


Figure 10.6 Principle of epitope tagging. An extra domain (an epitope tag, red) has been added genetically to one subunit (Rpb3) of the yeast RNA polymerase II. All the other subunits are normal, and assemble with the altered Rpb3 subunit to form an active polymerase. This polymerase has also been labeled by growing cells in labeled amino acids. **(a)** Add an antibody directed against the epitope tag, which immunoprecipitates the whole RNA polymerase, separating it from contaminating proteins (gray). This gives very pure polymerase in just one step. **(b)** Add the strong detergent SDS, which separates and denatures the subunits of the purified polymerase. **(c)** Electrophorese the denatured subunits of the polymerase to yield the electropherogram at bottom.

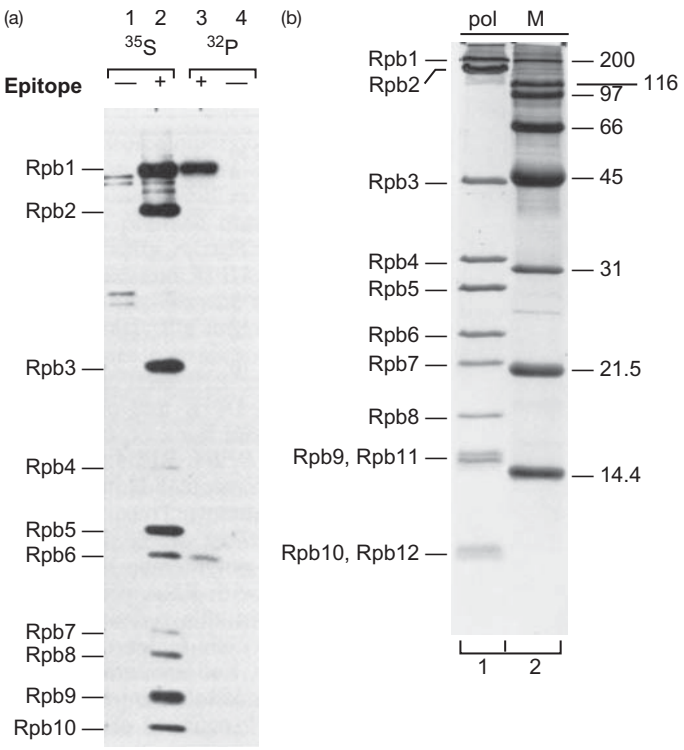


Figure 10.7 Subunit structure of yeast RNA polymerase II. **(a)** Apparent 10-subunit structure obtained by epitope tagging. Young and colleagues endowed one of the subunits of yeast polymerase II (Rpb3) with an extra group of amino acids (an epitope tag) by substituting a gene including the codons for this tag for the usual yeast *RPB3* gene. Then they labeled these engineered yeast cells with either [^{35}S]methionine to label all the polymerase subunits, or [γ-^{32}P]ATP to label the phosphorylated subunits only. They immunoprecipitated the labeled protein with an antibody directed against the epitope tag and electrophoresed the products. Lane 1, ^{35}S-labeled protein from wild-type yeast without the epitope tag; lane 2, ^{35}S-labeled protein from yeast having the epitope tag on Rpb3; lane 3, ^{32}P-labeled protein from yeast with the epitope tag; lane 4, ^{32}P-labeled protein from wild-type yeast. The polymerase II subunits are identified at left. **(b)** Apparent 12-subunit structure obtained by multistep purification including immunoprecipitation. Kornberg and colleagues immunoprecipitated yeast RNA polymerase II and subjected it to SDS-PAGE (lane 1), alongside molecular mass markers (lane 2). The marker molecular masses are given at right, and the polymerase II subunits are identified at left. Notice that Rpb9 and Rpb11 almost comigrate, as do Rpb10 and Rpb12. (*Sources:* (*a*) Kolodziej, P.A., N. Woychik, S.-M. Liao, and R. Young, RNA polymerase II subunit composition, stoichiometry, and phosphorylation, *Molecular and Cellular Biology* 10 (May 1990) p. 1917, f. 2. American Society for Microbiology. (*b*) Sayre, M.H., H. Tschochner, and R.D. Kornberg, Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry.* 267 (15 Nov 1992) p. 23379, f. 3b. American Society for Biochemistry and Molecular Biology.)

polymerase II subunits are phosphorylated, because they were labeled by [γ-^{32}P]ATP. These phosphoproteins are subunits Rpb1 and Rpb6. Rpb2 is also phosphorylated, but at such a low level that Figure 10.7a does not show it.

Core Subunits These three polypeptides, Rpb1, Rpb2, and Rpb3, are all absolutely required for enzyme activity. They are homologous to the β'-, β-, and α-subunits, respectively, of *E. coli* RNA polymerase.

How about functional relationships? We have seen (Chapter 6) that the *E. coli* β'-subunit binds DNA, and so does Rpb1. Chapter 6 also showed that the *E. coli* β-subunit is at or near the nucleotide-joining active site of the enzyme. Using the same experimental design, André Sentenac and his colleagues have established that Rpb2 is also at or near the active site of RNA polymerase II. The functional similarity among the second largest subunits in all three nuclear RNA polymerases, as well as prokaryotic polymerases, is mirrored by structural similarities among these same subunits, as revealed by the sequences of their genes.

Although Rpb3 does not closely resemble the *E. coli* α-subunit, there is one 20-amino-acid region of great similarity. In addition, the two subunits are about the same size and have the same stoichiometry, two monomers per holoenzyme. Furthermore, the same kinds of polymerase assembly defects are seen in *RPB3* mutants as in *E. coli* α-subunit mutants. All of these factors suggest that Rpb3 and *E. coli* α are homologous.

Common Subunits Five subunits—Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12—are found in all three yeast nuclear polymerases. We know little about the functions of these subunits, but the fact that they are found in all three polymerases suggests that they play roles fundamental to the transcription process.

SUMMARY The genes encoding all 12 RNA polymerase II subunits in yeast have been sequenced and subjected to mutation analysis. Three of the subunits resemble the core subunits of bacterial RNA polymerases in both structure and function, five are found in all three nuclear RNA polymerases, two are not required for activity, at least at 37°C, and two fall into none of these three categories. Two subunits, especially Rpb1, are heavily phosphorylated, and one is lightly phosphorylated.

Heterogeneity of the Rpb1 Subunit The very earliest studies on RNA polymerase II structure showed some heterogeneity in the largest subunit. Figure 10.8 illustrates this phenomenon in polymerase II from a mouse tumor called a plasmacytoma. We see three polypeptides near the top of the electrophoretic gel, labeled IIo, IIa, and IIb, that are present in smaller quantities than polypeptide IIc. These three polypeptides appear to be related to one another, and indeed two of them seem to derive from the other one. But which is the parent and which are the offspring? Sequencing of the yeast *RPB1* gene predicts a polypeptide product of 210 kD, so the IIa subunit, which has a molecular mass close to 210 kD, seems to be the parent.

Table 10.3 Yeast RNA Polymerase II Subunits

Subunit	SDS-PAGE Mobility (kD)	Protein Mass (kD)	Stoichiometry	Deletion Phenotype
Rpb1	220	190	1.1	Inviable
Rpb2	150	140	1.0	Inviable
Rpb3	45	35	2.1	Inviable
Rpb4	32	25	0.5	Conditional
Rpb5	27	25	2.0	Inviable
Rpb6	23	18	0.9	Inviable
Rpb7	17	19	0.5	Inviable
Rpb8	14	17	0.8	Inviable
Rpb9	13	14	2.0	Conditional
Rpb10	10	8.3	0.9	Inviable
Rpb11	13	14	1.0	Inviable
Rpb12	10	7.7	1.0	Inviable

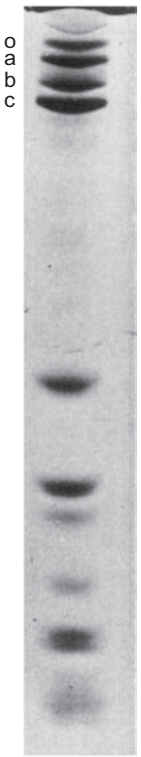


Figure 10.8 Partial subunit structure of mouse plasmacytoma RNA polymerase II. The largest subunits are identified by letter on the left, although these subunit designations are not the same as those applied to the yeast polymerase II (see Figure 10.7). Subunits o, a, and b are three forms of the largest subunit, corresponding to yeast Rpb1. Subunit c corresponds to yeast Rpb2. (*Source:* Sklar, V.E.F., L.B. Schwartz, and R.G. Roeder, Distinct molecular structures of nuclear class I, II, and III DNA-dependent RNA polymerases. *Proceedings of the National Academy of Sciences USA* 72 (Jan 1975) p. 350, f. 2C.)

Furthermore, amino acid sequencing has shown that the IIb subunit lacks a repeating string of seven amino acids (a *heptad*) with the following consensus sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser. Because this sequence is found at the carboxyl terminus of the IIa subunit, it is called the **carboxyl-terminal domain,** or **CTD.** Antibodies against the CTD react readily with the IIa subunit, but not with IIb, reinforcing the conclusion that IIb lacks this domain. A likely explanation for this heterogeneity is that a proteolytic enzyme clips off the CTD, converting IIa to IIb. Because IIb has not been observed in vivo, this clipping seems to be an artifact that occurs during purification of the enzyme. In fact, the sequence of the CTD suggests that it will not fold into a compact structure; instead, it is probably extended and therefore highly accessible to proteolytic enzymes.

What about the IIo subunit? It appears bigger than IIa, so it cannot arise through proteolysis. Instead, it seems to be a phosphorylated version of IIa. Indeed, subunit IIo can be converted to IIa by incubating it with a phosphatase that removes the phosphate groups. Furthermore, serines 2, 5, and sometimes 7 in the heptad are found to be phosphorylated in the IIo subunit.

Can we account for the difference in apparent molecular mass between IIo and IIa simply on the basis of phosphate groups? Apparently not; even though mammalian polymerase II contains 52 repeats of the heptad, not enough phosphates are present, so we must devise another explanation for the low electrophoretic mobility of IIo. Perhaps phosphorylation of the CTD induces a conformational change in IIo that makes it electrophorese more slowly and therefore seem larger than it really is. But this conformational change would have to persist even in the denatured protein. Figure 10.9 shows the probable relationships among the subunits IIo, IIa, and IIb.

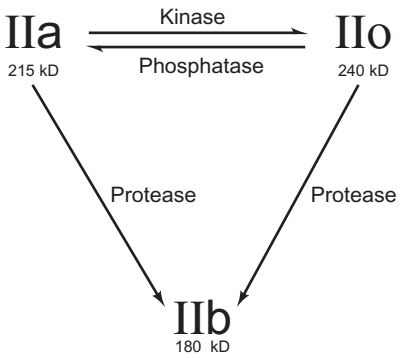


Figure 10.9 Proposed relationships among the different forms of the largest subunit of RNA polymerase II.

The fact that cells contain two forms of the Rpb1 subunit (IIo and IIa) implies that two different forms of RNA polymerase II exist, each of which contains one of these subunits. We call these **RNA polymerase IIO** and **RNA polymerase IIA,** respectively. The nonphysiological form of the enzyme, which contains subunit IIb, is called RNA polymerase IIB.

Do polymerases IIO and IIA have identical or distinct roles in the cell? The evidence strongly suggests that IIA (the unphosphorylated form of the enzyme) is the species that initially binds to the promoter, and that IIO (with its CTD phosphorylated) is the species that carries out elongation. Thus, phosphorylation of the CTD appears to accompany the transition from initiation to elongation. We will examine the evidence for this hypothesis, and refine it further, in Chapter 11.

SUMMARY Subunit IIa is the primary product of the *RPB1* gene in yeast. It can be converted to IIb in vitro by proteolytic removal of the carboxyl-terminal domain (CTD), which is essentially a heptapeptide repeated over and over. Subunit IIa can be converted to IIo by phosphorylating two serines in the repeating heptad that makes up the CTD. The enzyme (polymerase IIA) with the IIa subunit is the one that binds to the promoter; the enzyme (polymerase IIO) with the IIo subunit is the one involved in transcript elongation.

The Three-Dimensional Structure of RNA Polymerase II

The most powerful method for determining the shape of a protein, as we have seen in Chapter 9, is x-ray crystallography. This has been done with RNA polymerases from *Thermus aquaticus* and phage T7, but, until 1999, it was difficult to produce crystals of RNA polymerase II of high enough quality for x-ray crystallography studies. The problem lay in the heterogeneity of the polymerase caused by the loss of the Rpb4 and Rpb7 subunits from some of the enzymes. (Heterogeneous mixtures of proteins do not form crystals readily.) Roger Kornberg and colleagues solved this heterogeneity problem by using a mutant yeast polymerase (pol II Δ4/7) lacking Rbp4 (and therefore lacking Rpb7, because Rpb7 binds to Rpb4 and depends on the latter for binding to the rest of the enzyme). This polymerase is capable of transcription elongation, though not initiation at promoters. Thus, it should be adequate for modeling the elongation complex. It produced crystals that were good enough for x-ray crystallography leading to a model with up to 2.8 Å resolution in 2001.

Figure 10.10 presents a stereo view of this model of yeast RNA polymerase II. Each of the subunits is color-coded and their relative positions are illustrated in the small diagram at the upper right. The most prominent feature of the enzyme is the deep DNA-binding cleft, with the active site, containing a Mg^{2+} ion, at the base of the cleft. The opening of the cleft features a pair of jaws. The upper jaw is composed of part of Rpb1 plus Rpb9, and the lower jaw is composed of part of Rpb5.

Previous, lower resolution structural studies by Kornberg and colleagues had shown that the DNA template lay in the cleft in the enzyme. The newer structure strengthened this hypothesis by showing that the cleft is lined with basic amino acids, whereas almost the entire remainder of the surface of the enzyme is acidic. The basic residues in the cleft presumably help the enzyme bind to the acidic DNA template.

Structural studies of all single-subunit RNA and DNA polymerases had shown two metal ions at the active center, and a mechanism relying on both metal ions was therefore proposed. Thus, it came as a surprise to find only one Mg^{2+} ion in previous crystal structures of yeast polymerase II. However, the higher-resolution structure showed two Mg^{2+} ions, though the signal for one of them was weak. Kornberg and colleagues theorized that the strong metal signal corresponds to a strongly bound Mg^{2+} ion *(metal A),* but the weak signal corresponds to a weakly bound Mg^{2+} ion *(metal B)* that may enter bound to the substrate nucleotide. Metal A is bound to three invariant aspartate residues (D481, D483, and D485 of Rpb1). Metal B is also surrounded by three acidic residues (D481 of Rpb1 and E836 and D837 of Rpb2), but they are too far away in the crystal structure to coordinate the metal. Nevertheless, during catalysis, they may move closer to metal B, coordinate it, and thereby create the proper conformation at the active center to accelerate the polymerase reaction.

SUMMARY The structure of yeast polymerase II (pol II Δ4/7) reveals a deep cleft that can accept a DNA template. The catalytic center, containing a Mg^{2+} ion, lies at the bottom of the cleft. A second Mg^{2+} ion is present in low concentration, and presumably enters the enzyme bound to each substrate nucleotide.

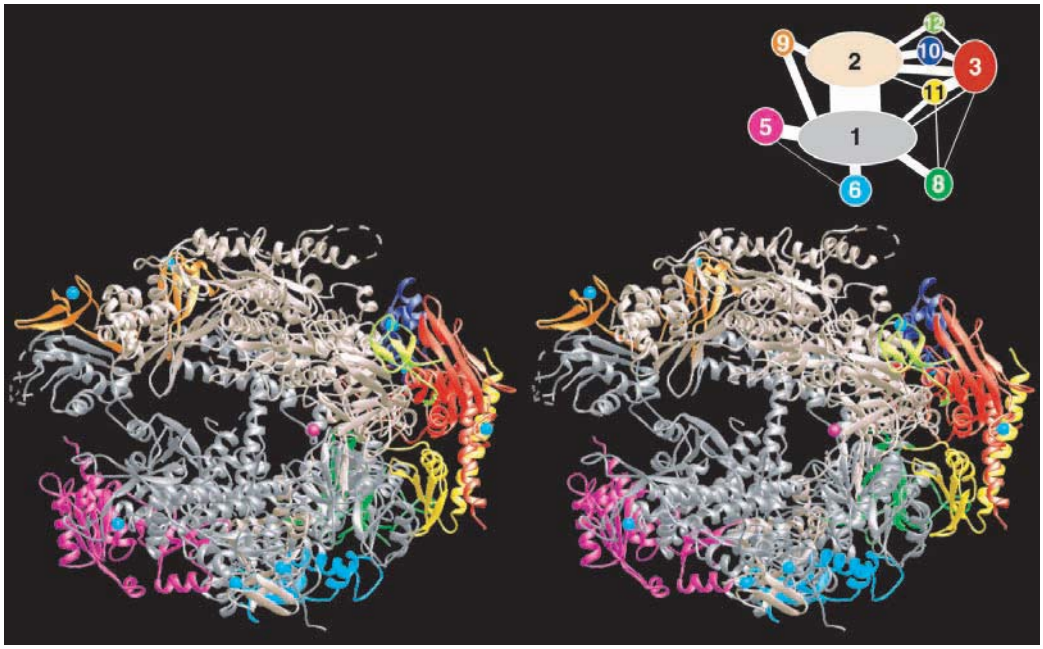


Figure 10.10 Crystal structure of yeast RNA polymerase II. The stereo view at bottom shows all 10 subunits of the enzyme (lacking Rpb4 and Rpb7), color-coded according to the small diagram at the upper right. The thickness of the white lines connecting the subunits in the small diagram indicate the extent of contact between the subunits. The metal ion at the active center in the stereo view is represented by a magenta sphere. Zn^{2+} ions are represented by blue spheres. (*Source:* Cramer, et al., *Science* 292: p. 1864.)

Three-Dimensional Structure of RNA Polymerase II in an Elongation Complex The previous section has shown the shape of yeast RNA polymerase II by itself. But Kornberg and colleagues have also determined the structure of yeast polymerase II bound to its DNA template and RNA product in an elongation complex. The resolution is not as high (3.3 Å) as in the structure of the polymerase by itself, but it still gives a wealth of information about the interaction between the enzyme and the DNA template and RNA product.

To induce polymerase II to initiate on its own without help from any transcription factors, Kornberg and colleagues used a DNA template with a 3′-single-stranded oligo[dC] tail, which allows polymerase II to initiate in the tail, 2–3nt from the beginning of the double-stranded region. The template was also designed to allow the polymerase to elongate the RNA to a 14-mer in the absence of UTP and then pause at the point where it needed the first UTP. This sequence of events created a homogeneous population of elongation complexes, contaminated with inactive polymerases that did not bind to DNA. The inactive enzymes were removed on a heparin column. Heparin is a polyanionic substance that can bind in the basic cleft of the polymerase if the cleft is not occupied by DNA. Thus, inactive enzymes bound to the heparin on the column, but the active elongation complexes passed through. These complexes could then be crystallized.

Figure 10.11a shows the crystal structure of the elongation complex, together with the crystal structure of the polymerase by itself. One of the most obvious differences, aside from the presence of the nucleic acids in the elongation complex, is the position of the **clamp.** In the polymerase itself, the clamp is open to allow access to the active site. But in the elongation complex, the clamp is closed over the DNA template and RNA product. This ensures that the enzyme will be **processive**—able to transcribe a whole gene without falling off and terminating transcription prematurely.

Figure 10.11b shows a closer view of the elongation complex, with part of the enzyme cut away to reveal the nucleic acids in the enzyme's cleft. Several features are apparent. We can see that the axis of the DNA–RNA hybrid (formed from the template DNA strand and the RNA product) lies at an angle with respect to the downstream DNA duplex that has yet to be transcribed. This turn is forced by the closing of the clamp and is facilitated by the single-stranded DNA between the RNA–DNA hybrid and the downstream DNA duplex. (Kornberg and colleagues' later crystal structure of a post-translocation complex showed that the RNA–DNA hybrid is actually 8 bp long.)

We can also see the catalytic Mg^{2+} ion at the active center—the point where a nucleotide has just been added to the growing RNA chain. This ion corresponds to metal A detected in the structure of polymerase itself. Finally, we can see a **bridge helix** that spans the cleft near the active center. We will discuss this bridge helix in more detail later in this section.

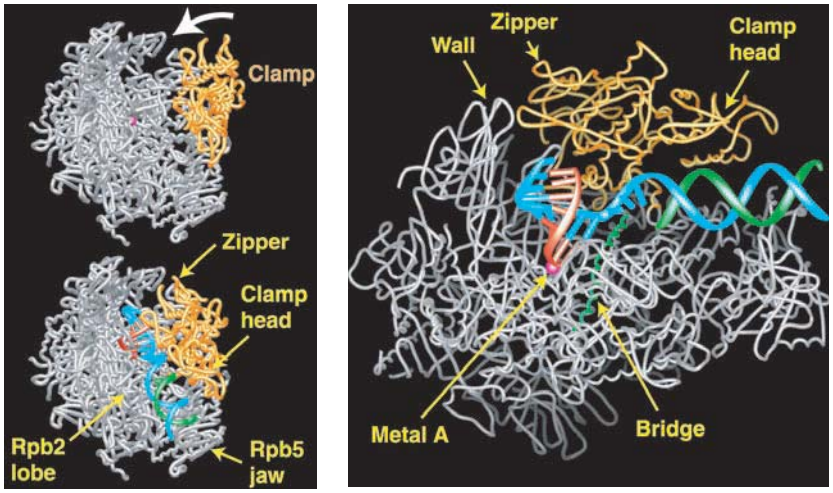


Figure 10.11 Crystal structure of the elongation complex. **(a)** Comparison of the crystal structures of the free polymerase II (top) and the elongation complex (bottom). The clamp is highlighted in yellow. The template DNA strand, the nontemplate DNA strand, and RNA product are highlighted in blue, green, and red, respectively. **(b)** Detailed view of the elongation complex. Color codes are the same as in panel (a). The active center metal is in magenta and the bridge helix is in green. (*Source:* Gnatt et al., *Science* 292: p. 1877.)

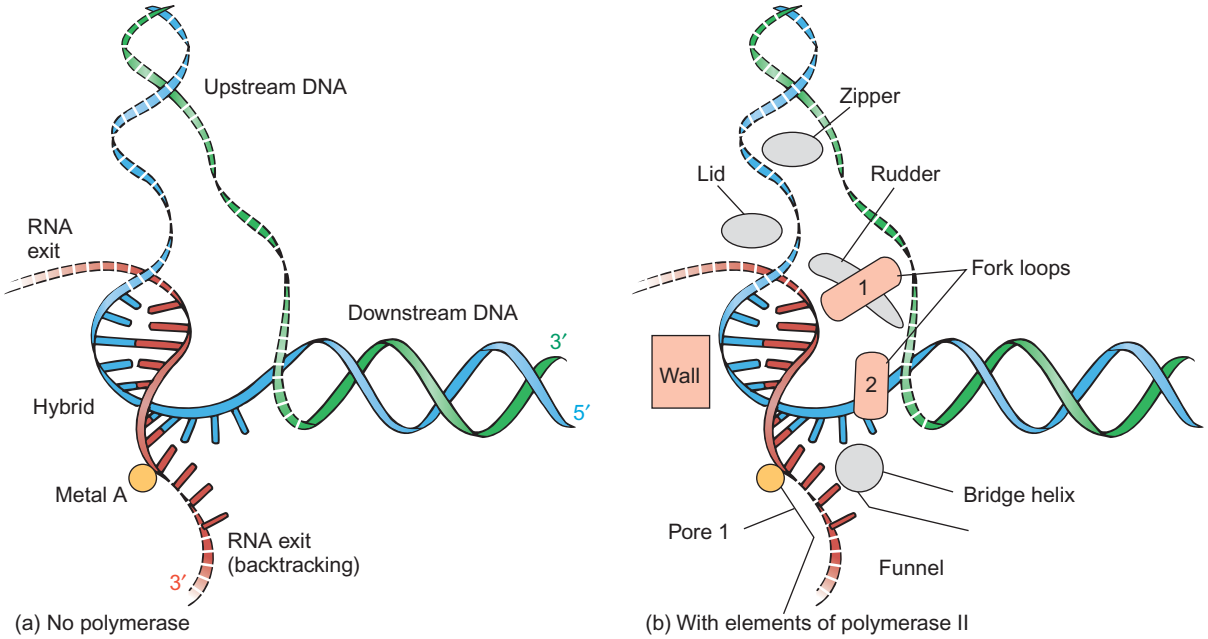


Figure 10.12 The transcription bubble. **(a)** Positions of the nucleic acids. The DNA template strand is in blue, the nontemplate strand in green, and the RNA in red. Solid lines correspond to nucleic acids represented in the crystal structure. Dashed lines show hypothetical paths for nucleic acids not represented in the crystal structure. **(b)** Nucleic acids plus key elements of RNA polymerase II. The nucleic acids from panel (a) are superimposed on critical elements of polymerase II: the protein loops extending from the clamp (the zipper, lid, and rudder); fork loops 1 and 2; the bridge helix; the funnel; pore 1; and the wall. (*Source:* Adapted from Gnatt, A.L., P. Cramer, J. Fu, D.A. Bushnell, and R.D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292 (2001) p. 1879, f. 4.)

The Mg^{2+} ion in the elongation complex (metal A) is positioned so that it can bind to the phosphate linking nucleotides +1 and −1 (the last two nucleotides added to the growing RNA; Figure 10.12a). Metal B is missing from this complex, presumably because it has departed along with the pyrophosphate released from the last nucleotide added to the RNA. The nucleotide in position +1 lies just at the entrance to **pore 1** (Figure 10.12b), strongly suggesting that the nucleotides enter the active site through this pore. Indeed, there would not be room for them to enter any other way without significant rearrangements of the nucleic acids and proteins. Moreover,

pore 1 is in perfect position for extrusion of the 3′-end of the RNA when the polymerase backtracks. Such backtracks occur when a nucleotide is misincorporated (recall Chapter 6), thus exposing the misincorporated nucleotide to removal by TFIIS (Chapter 11), which binds to the funnel at the other end of the pore 1.

Figure 10.12b also illustrates the probable roles of three loops, called the *lid*, *rudder*, and *zipper*, which extend from the clamp. These loops are in position to affect several important events, including formation and maintenance of the transcription bubble and dissociation of the RNA–DNA hybrid. If the RNA–DNA hybrid extended farther than 9 bp, the rudder would be in the way. Thus, the rudder may facilitate the dissociation of the hybrid.

Kornberg and colleagues noted that the bridge helix is straight in the elongation complex, but bent in the bacterial polymerase crystal structures. This bend occurs in the neighborhood of conserved residues corresponding to Thr 831 and Ala 832 and would interfere with nucleotide binding to the active site. This observation led these authors to speculate about the role of the bridge helix in **translocation** (the 1-nt steps of DNA template and RNA product through the polymerase), as illustrated in Figure 10.13. They suggest that the bridge helix oscillates between straight and bent conformations during the translocation step as follows: With the bridge helix in the straight state, the active site is open for addition of a nucleotide, so the nucleotide enters through pore 1 of the enzyme, just below the active site. The polymerase adds this new nucleotide to the growing RNA chain, filling the space between the 3′-end of the RNA and the straight bridge helix. Next, coincident with translocation, the bridge helix shifts to the bent state. When it shifts back to the straight state, it reopens the space at the 3′-end of the RNA, and the cycle is ready to repeat.

Further support for this hypothesis comes from the crystal structure of the cocrystal of yeast RNA polymerase II and α-amanitin. The α-amanitin-binding site lies so close to the bridge helix that hydrogen bonds form between the two. Binding of α-amanitin to this site thus severely constrains the bending of the bridge helix necessary for translocation. This explains how α-amanitin can block RNA synthesis without blocking nucleotide entry or phosphodiester bond formation—it blocks translocation after a phosphodiester bond forms.

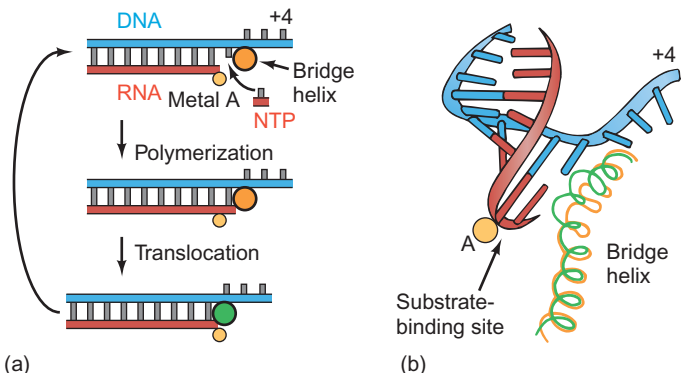


Figure 10.13 Proposed translocation mechanism. (a) The model. We begin with the bridge helix in the straight state (orange), leaving a gap for a nucleotide (NTP) to enter the active site, marked by the yellow circle (metal A). During the synthesis step, the nucleotide joins the growing RNA (red), filling the gap between the end of the RNA and the straight bridge helix. During the translocation step, the RNA–DNA hybrid moves one bp to the left, bringing a new template strand nucleotide into the active site. Simultaneously, the bridge helix bends (represented by the green dot), remaining close to the end of the RNA. When the bridge helix returns to the straight state (arrow at left), it reopens the active site so another nucleotide can enter. **(b)** The straight and bent states of the bridge helix. The straight state is represented by the orange helix, and the bent state by the green helix. Note that bending the bridge helix brings it very close to the end of the growing RNA. (*Source:* Adapted from Gnatt, A.L., P. Cramer, J. Fu, D.A. Bushnell, and R.D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292 (2001) p.1880, F.6.)

SUMMARY The crystal structure of a transcription elongation complex involving yeast RNA polymerase II (lacking Rpb 4/7) reveals that the clamp is indeed closed over the RNA–DNA hybrid in the enzyme's cleft, ensuring processivity of transcription. In addition, three loops of the clamp—the rudder, lid, and zipper—appear to play important roles in, respectively: initiating dissociation of the RNA–DNA hybrid, maintaining this dissociation, and maintaining dissociation of the template DNA. The active center of the enzyme lies at the end of pore 1, which appears to be the conduit for nucleotides to enter the enzyme and for extruded RNA to exit the enzyme during backtracking. A bridge helix lies adjacent to the active center, and flexing of this helix could play a role in translocation during transcription. Binding of α-amanitin to a site near this helix appears to block flexing of the helix, and therefore blocks translocation.

Structural Basis of Nucleotide Selection In 2004, Kornberg and colleagues published x-ray diffraction data on a posttranslocation complex. First, they bound RNA polymerase II to a set of synthetic oligonucleotides representing a partially double-stranded DNA template and a 10-nt RNA product terminated in 3′-deoxyadenosine, which, as we have just seen, prevents addition of any more nucleotides, and traps the polymerase in the posttranslocation state. Then they soaked crystals of this complex with either a nucleotide (UTP) that paired correctly with the next nucleotide in the DNA template strand, or a mismatched nucleotide, then obtained the crystal structures of the resulting complexes. The difference between the two structures was striking: The mismatched nucleotide lay in a site adjacent to the one occupied by the correct nucleotide, and it was inverted relative to the correct nucleotide (Figure 10.14).

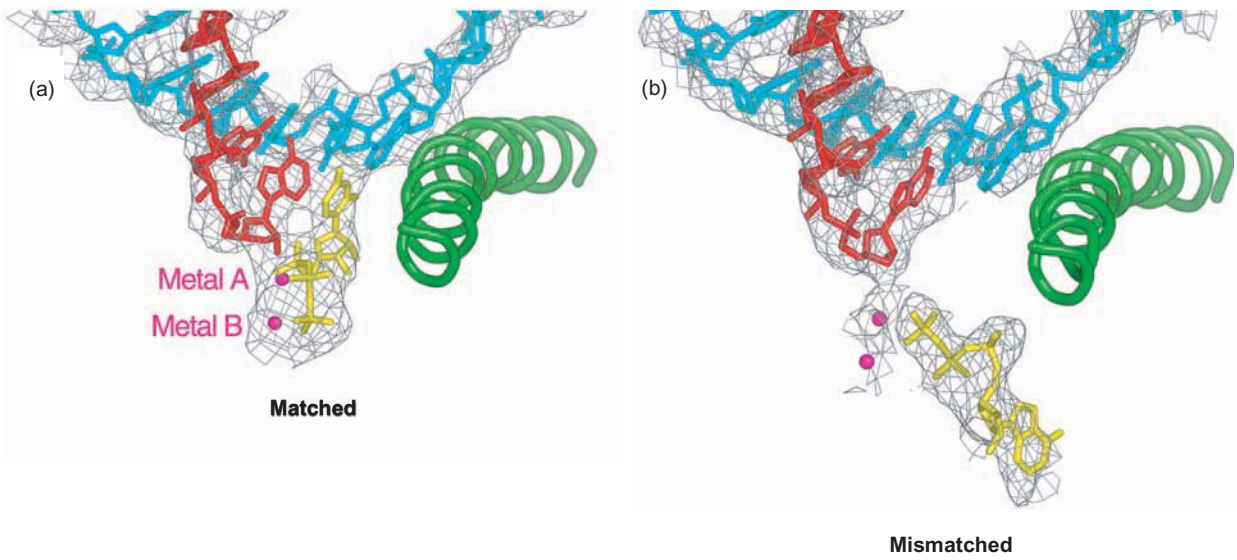


Figure 10.14 Matched (a) and mismatched (b) nucleotides in A and E sites, respectively. Metals A and B at the active site are labeled and represented by magenta spheres. DNA is in blue, RNA is in red, and the nucleotides in the A and E sites are in yellow. The green coil is the bridge helix of the RNA polymerase. (*Source:* Reprinted from Cell, Vol. 119, Kenneth D. Westover, David A. Bushnell and Roger D. Kornberg, "Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center," p. 481–489, Copyright 2004 with permission from Elsevier.

These data revealed two distinct nucleotide-binding sites at the active center of RNA polymerase II. The previously-known site, where phosphodiester bond formation, or nucleotide addition, occurs, had already been named the **A site, for "addition."** The second site, where nucleotides bind prior to entering the A site, had been predicted by Alexander Goldfarb and colleagues based on biochemical studies of the *E. coli* RNA polymerase; they had named this the **E site,** for "entry." The two sites overlap somewhat and Kornberg and colleagues noted that nucleotides, in moving through the nucleotide entry pore toward the A site, must pass through the E site.

The crystal structures also reinforced the case for two metal ions at the active site. One metal ion (metal A) is permanently attached to the enzyme, but the other (metal B) enters the enzyme attached to the incoming nucleotide (coordinated to the β- and γ-phosphates). In contrast to previous structures, the two metal ions had equivalent intensities in the latest structures. Thus, the mechanism of phosphodiester bond formation in RNA polymerases almost certainly relies on two metal ions at the active site.

The discovery of the E and A sites, though interesting, did not illuminate the mechanism by which the polymerase discriminates among the four ribonucleoside triphosphates, or how it excludes dNTPs. Then, in 2006, Kornberg and colleagues obtained the crystal structure of a very similar complex, but with GTP, rather than UTP, in the A site, opposite a C, rather than an A, in the template i+1 site. In this structure, and in a further refined version of their previous structure, they could see the **trigger loop,** a part of Rpb1 roughly encompassing residues 1070 to 1100, very near the substrate in the A site (Figure 10.15a). In both of these structures, the correct nucleotide occupied the A site. In 12 other crystal structures without the correct substrate in the A site, three alternative positions for the trigger loop were observed, all remote from the A site (Figure 10.15b).

Thus, only when the correct substrate nucleotide occupies the A site does the trigger loop come into play, and then it makes several important contacts with the substrate. These contacts presumably stabilize the substrate's association with the active site, and thereby contribute to the specificity of the enzyme. Indeed, as Figure 10.16a shows, the trigger loop is involved in a network of interactions involving the substrate (GTP in this case), the bridge helix, and other amino acids of Rpb1 and Rpb2 at the active site. For example, Leu 1081 makes a hydrophobic contact with the substrate base, and Gln 1078 engages in a hydrogen bond network with Rpb1-Asn 479 and the 3′-hydroxyl group of the substrate ribose. Indeed, there could even be a weak direct H-bond between this 3′-hydroxyl group and Gln 1078. In addition, His 1085 makes an H-bond or salt bridge to the β-phosphate of the substrate, and His 1085 is held in proper position by H-bonds to Asn 1082 and the Rpb2-Ser1019 backbone carbonyl group. Finally, Rpb1 Arg 446 (not part of the trigger loop) lies close to the 2′-hydroxyl group of the substrate ribose. Thus, this network of contacts recognizes all parts of the substrate nucleotide: the base, both hydroxyl groups of the sugar, and one of the phosphates.

Why is this network of contacts so important to nucleotide specificity? Presumably, the enzyme requires these contacts to create the proper environment for catalysis. Even more explicitly, the trigger loop His 1085

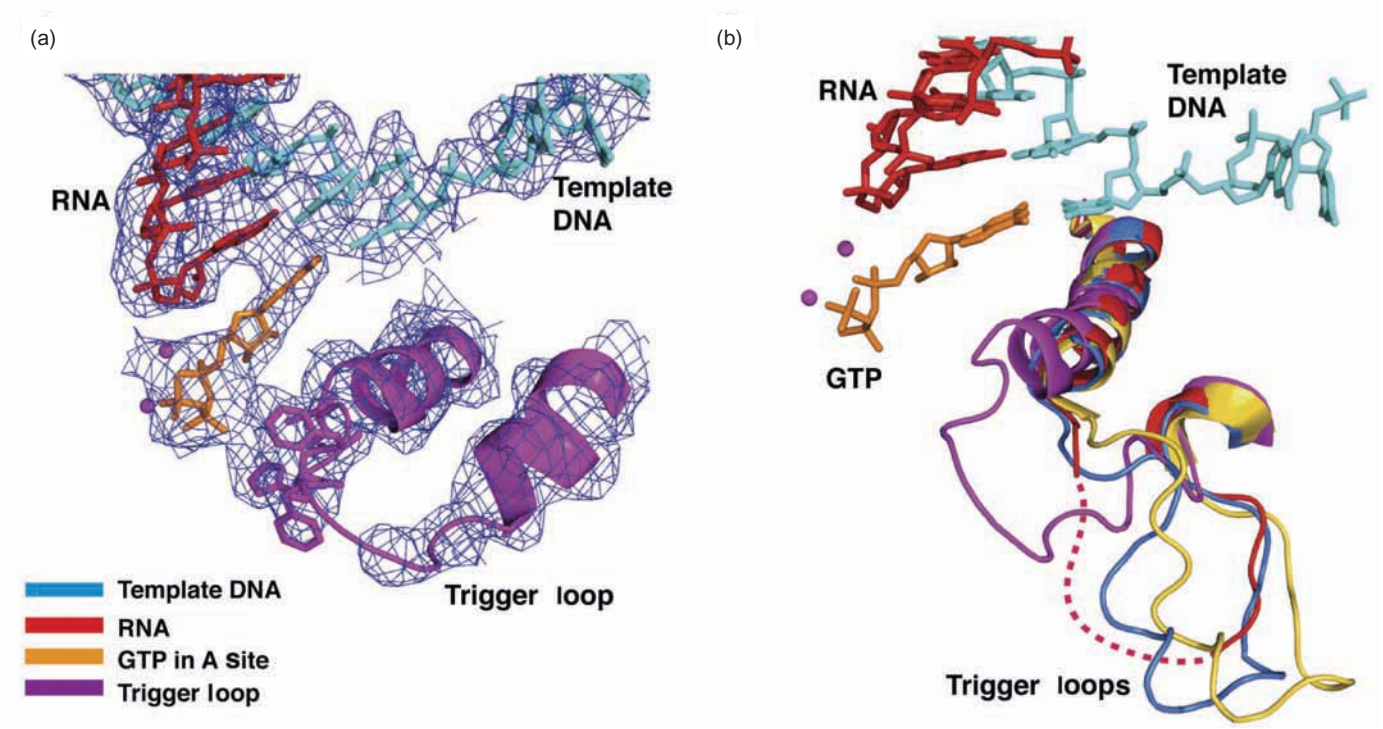


Figure 10.15 RNA polymerase II active site, including trigger loop. (a) The active site is shown with the proper NTP (GTP) in the A site. The electron densities are modeled with blue mesh. The trigger loop is in magenta, the GTP in orange, the RNA in red, and the template DNA strand in cyan. The Mg^{2+} ions are represented by magenta spheres. **(b)** Four different conformations for the trigger loop. Magenta, as in panel (a), with GTP in the A site at low Mg^{2+} concentration; red, ATP in the E site, low Mg^{2+}; blue, UTP in the E site, high Mg^{2+}; yellow, RNA polymerase II-TFIIS complex (see Chapter 11) with no nucleotide and high Mg^{2+}. (*Source:* Reprinted from CELL, Vol. 127, Wang et al, Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis, Issue 5, 1 December 2006, pages 941–954, © 2006, with permission from Elsevier.)

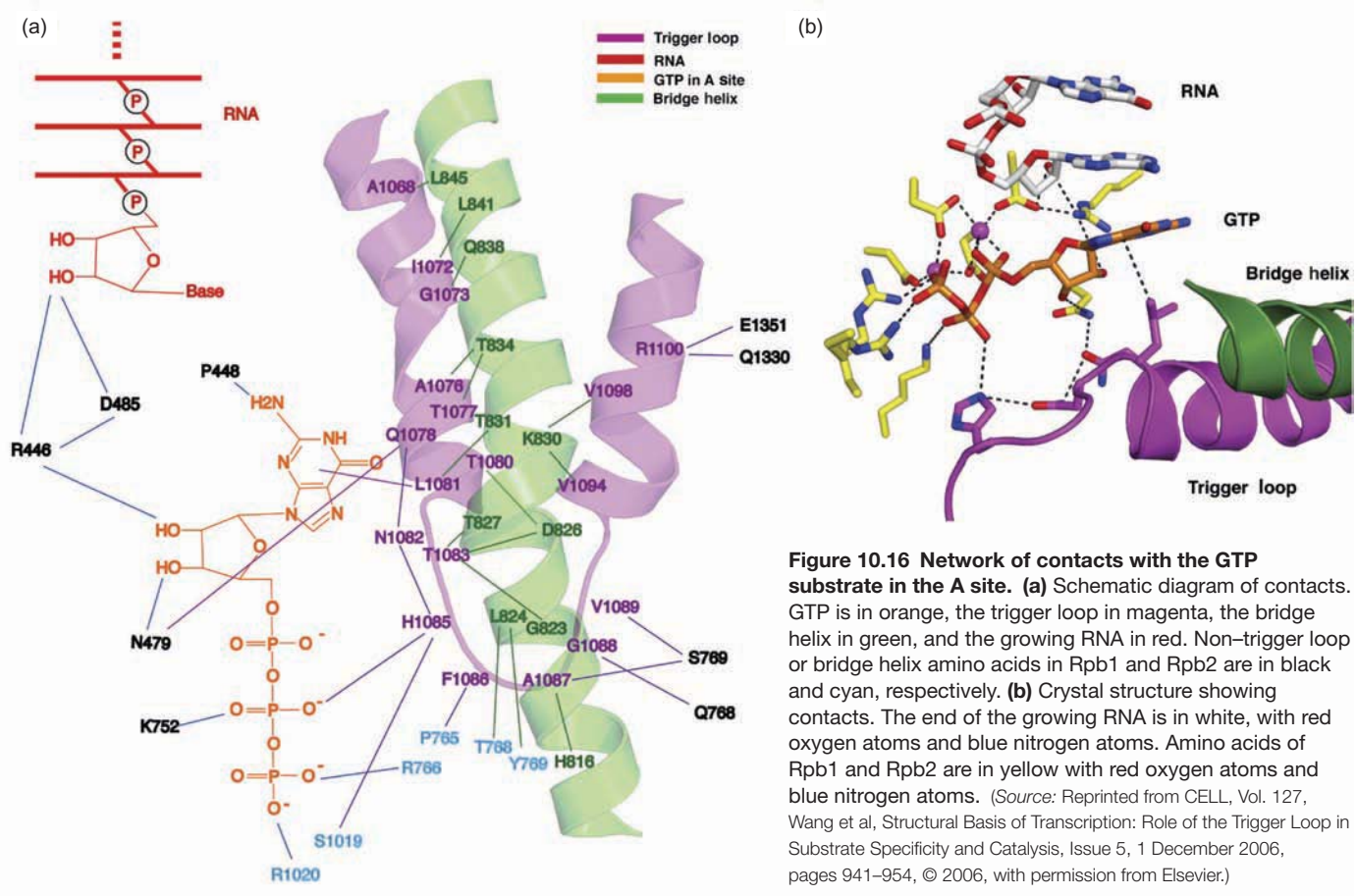


Figure 10.16 Network of contacts with the GTP substrate in the A site. (a) Schematic diagram of contacts. GTP is in orange, the trigger loop in magenta, the bridge helix in green, and the growing RNA in red. Non–trigger loop or bridge helix amino acids in Rpb1 and Rpb2 are in black and cyan, respectively. **(b)** Crystal structure showing contacts. The end of the growing RNA is in white, with red oxygen atoms and blue nitrogen atoms. Amino acids of Rpb1 and Rpb2 are in yellow with red oxygen atoms and blue nitrogen atoms. (*Source:* Reprinted from CELL, Vol. 127, Wang et al, Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis, Issue 5, 1 December 2006, pages 941–954, © 2006, with permission from Elsevier.)

contact with the β-phosphate of the substrate may have catalytic implications. The histidine imidazole group is protonated at physiological pH and would therefore be expected to withdraw negative charge from the β-phosphate, which could in turn decrease the negativity of the γ-phosphate. Because the γ-phosphate is the target of a nucleophilic attack by the terminal 3′-hydroxyl group of the growing RNA, decreasing its negative charge should make it a better nucleophilic target and therefore help catalyze the reaction.

What about discrimination against dNTPs? Kornberg and colleagues found that they could prepare enzyme-substrate complexes with dNTPs in the A site, but that the enzyme incorporated deoxyribonucleotides at a much slower rate than it did ribonucleotides. They concluded that the enzyme makes this discrimination, not at the substrate binding step, but at the catalytic step. Moreover, the enzyme seems to have a way of removing a deoxyribonucleotide even after it has been incorporated. Figure 10.16a shows that Rpb1 Arg 446 and Glu 485 contact the 2′-hydroxyl group of the nucleotide that had been incorporated just before the new substrate bound. If this hydroxyl group is missing because a dNMP was incorporated by accident, these contacts can't be made, and the enzyme will presumably stall until the misincorporated dNMP can be removed.

SUMMARY In moving through the entry pore toward the active site of RNA polymerase II, an incoming nucleotide first encounters the E (entry) site, where it is inverted relative to its position in the A site, the active site where phosphodiester bonds are formed. Two metal ions (Mg^{2+} or Mn^{2+}) are present at the active site. One is permanently bound to the enzyme and one enters the active site complexed to the incoming nucleotide. The trigger loop of Rpb1 positions the substrate for incorporation and discriminates against improper nucleotides.

The Role of Rpb4 and Rpb7 The studies we have been discussing were very informative, but they told us nothing about the role of Rpb4 and Rpb7, because these two subunits were missing from the core polymerase II that Kornberg and colleagues crystallized. To fill in this gap, two groups, one led by Patrick Cramer, and the other by Kornberg, succeeded in crystallizing the complete, 12-subunit enzyme from yeast. Cramer's group solved the problem of producing a homogeneous population of 12-subunit enzyme by incubating the purified 10-subunit enzyme with an excess of Rbp4/7 produced in *E. coli* from cloned genes. Kornberg's group purified the 12-subunit enzyme directly by affinity chromatography, using an antibody directed against an epitope tag added to the Rpb4 subunit. They further enhanced their chances of isolating the intact enzyme by isolating the enzyme from stationary phase yeast cells, which contain a high proportion of 12-subunit enzyme, rather than the 10-subunit core enzyme.

Figure 10.17 shows the crystal structure that Cramer and colleagues obtained for the 12-subunit enzyme. The subunits Rbp4 and Rpb7 are immediately apparent because they stick out to the side of the enzyme, rather like a wedge, with its thin end lodged in the rest of the polymerase (the core enzyme). Furthermore, Cramer and colleagues noticed that the presence or absence of Rpb4/7 determines the position of the clamp of the enzyme. Without Rpb4/7, the clamp is free to swing open, but, as the inset at the lower right in Figure 10.17a shows, when wedge-like Rpb4/7 is present, the wedge forces the clamp shut.

What does this new information tell us about how the polymerase associates with promoter DNA? Cramer and colleagues, as well as Bushnell and Kornberg, suggested that the polymerase core could bind to the promoter in double-stranded form, the promoter could then melt, and then Rpb4/7 could bind and close the clamp over the template DNA strand, excluding the nontemplate strand from the active site. But these authors also point out that this simple model is contradicted by other evidence: First, RNA polymerases from other organisms have Rpb4/7 homologs that are not thought to dissociate from the core enzyme. Similarly, the crystal structure of the *E. coli* RNA polymerase holoenzyme, the form of the enzyme involved in initiation (Chapter 6), has a closed conformation that seems incapable of allowing access to double-stranded DNA. So both sets of authors proposed that the promoter DNA could bind to the outer surface of the enzyme and melt, and the template strand could then descend into the active site, with accompanying pronounced bending of the promoter DNA.

Both research groups also noted a potential strong influence of Rpb4/7 on interaction with general transcription factors, which we will discuss in Chapter 11. We know that RNA polymerase II cannot bind to promoter DNA without help from several general transcription factors, and some of these make direct contact with an area of the polymerase called the "dock" region. Rpb4/7 greatly extends the dock region, as shown in Figure 10.17b. Thus, Rpb4/7 could play a major role in binding the vital general transcription factors.

Further work has shown that Rpb7 can bind to a nascent RNA. This finding, together with the proximity of Rpb4/7 to the base of the CTD of Rpb1 has prompted the suggestion that it can bind the nascent RNA and direct it toward the CTD. This could be important because, as we will see in Chapters 14 and 15, the CTD harbors proteins that make essential modifications (splicing, capping, and polyadenylation) to nascent mRNAs.

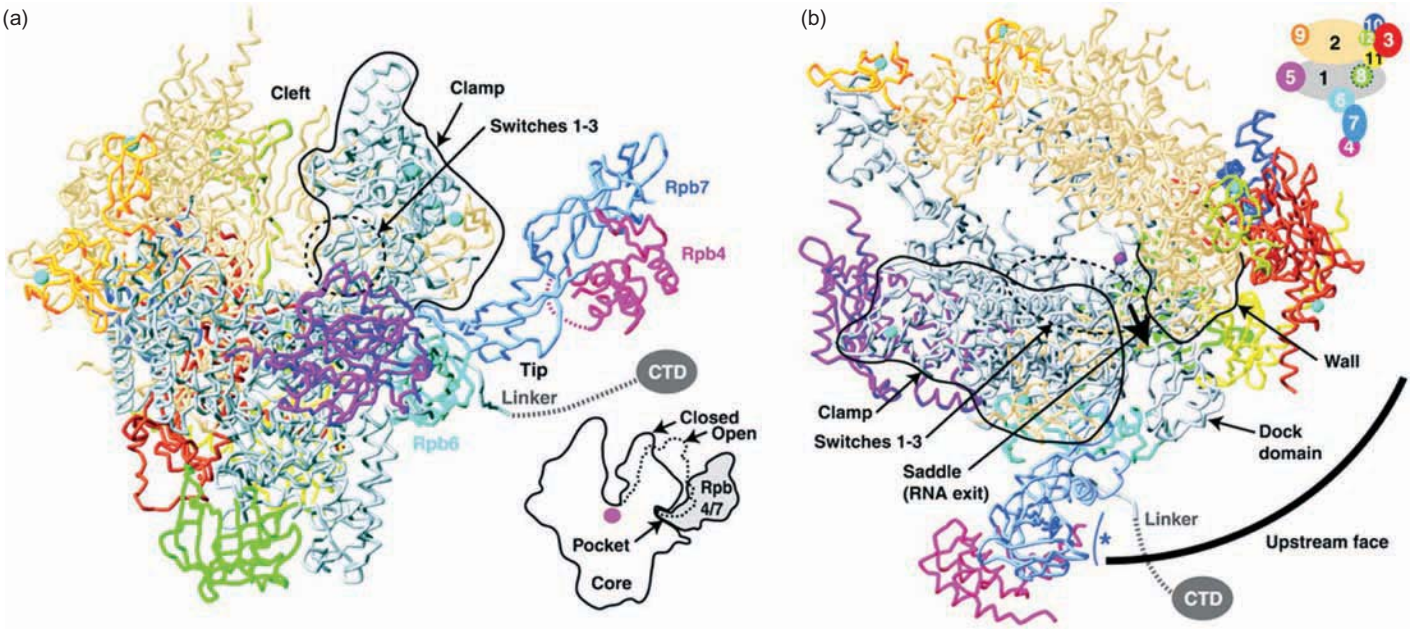


Figure 10.17 Crystal structure of the 12-subunit RNA polymerase II from yeast. (a) Structure showing the interaction between Rpb4/7 and the core polymerase. Rpb4 and Rpb7 are in magenta and blue, respectively, and are labeled. The clamp is outlined in solid black. The location of switches 1–3 is denoted by a dashed circle. Eight zinc ions are denoted by cyan spheres, and the magnesium ion at the active center at the base of the cleft (difficult to see in this panel) is represented by a pink sphere. The linker to the CTD of Rpb1 is denoted by a dashed line. The inset at lower right shows the closed and open positions of the clamp, and demonstrates that binding of Rpb4/7 is incompatible with the clamp's open position; that is, binding of Rpb4/7 wedges the clamp shut. **(b)** Another view of the structure, with the subunits color-coded as shown at upper right. This view emphasizes the effect of Rpb4/7 on extension of the dock domain of the enzyme. The solid circle segment at lower right represents a 25-bp radius, centered on the active site, which is the minimum distance between the TATA box and the transcription start site. The blue asterisk at lower center indicates a potential RNA-binding site on Rpb7. (*Source: (a-b)* © 2003 National Academy of Sciences Proceedings of the National Academy of Sciences, Vol. 100, no. 12, June 10, 2003, p. 6964–6968 "Architecture of initiation-competent 12-subunit RNA polymerase II," Karim-Jean Armache, Hubert Kettenberger, and Patrick Cramer, Fig. 2, p. 6966.

SUMMARY The structure of the 12-subunit RNA polymerase II reveals that, with Rpb4/7 in place, the clamp is forced shut. Because initiation occurs with the 12-subunit enzyme, with its clamp shut, it appears that the promoter DNA must melt before the template DNA strand can descend into the enzyme's active site. It also appears that Rpb4/7 extends the dock region of the polymerase, making it easier for certain general transcription factors to bind, thereby facilitating transcription initiation.

10.2 Promoters

We have seen that the three eukaryotic RNA polymerases have different structures and they transcribe different classes of genes. We would therefore expect that the three polymerases would recognize different promoters, and this expectation has been borne out. We will conclude this chapter by looking at the structures of the promoters recognized by all three polymerases.

Class II Promoters

We begin with the promoters recognized by RNA polymerase II (**class II promoters**) because these are the most complex and best studied. Class II promoters can be considered as having two parts: the **core promoter** and the **proximal promoter.** The core promoter attracts general transcription factors and RNA polymerase II at a basal level and sets the transcription start site and direction of transcription. It consists of elements lying within about 37 bp of the transcription start site, on either side. The proximal promoter helps attract general transcription factors and RNA polymerase and includes promoter elements that can extend from about 37 bp up to 250 bp upstream of the transcription start site. Elements of the proximal promoter are also sometimes called **upstream promoter elements.**

The core promoter is modular and can contain almost any combination of the following elements (Figure 10.18). The **TATA box** is centered at approximately position −28 (about −31 to −26) and has the consensus sequence TATA(A/T)AA(G/A); the **TFIIB recognition element (BRE)** lies just upstream of the TATA box (about

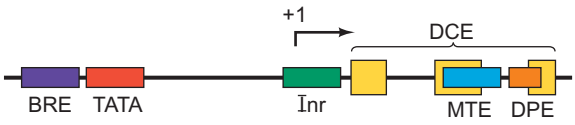


Figure 10.18 A generic class II core promoter. This core promoter contains up to six elements. These are, 5′ to 3′: the TFIIB-recognition element (BRE, purple); the TATA box (red); the initiator (green); the downstream core element, in three parts (DCE, yellow); the motif ten element (MTE, blue); and the downstream promoter element (DPE, orange). The exact locations of these promoter elements are given in the text.

position −37 to −32) and has the consensus sequence (G/C)(G/C)(G/A)CGCC; the **initiator** (**Inr**) is centered on the transcription start site (position −2 to +4) and has the consensus sequence GCA(G/T)T(T/C) in *Drosophila,* or PyPyAN(T/A)PyPy in mammals; the **downstream promoter element** (**DPE**) is centered on position +30 (+28 to +32); the **downstream core element** (**DCE**) has three parts located at approximately +6 to +12, +17 to +23, and +31 to +33, and these have the consensus sequences CTTC, CTGT, and AGC, respectively; and the **motif ten element** (**MTE**) lies approximately between positions +18 and +27.

The TATA Box By far the best-studied element in the many class II promoters is a sequence of bases with the consensus sequence TATAAA (in the nontemplate strand). The last A of this sequence usually lies 25 to 30 bp upstream of the transcription start site in higher eukaryotes. Its name, *TATA box,* derives from its first four bases. You may have noticed the close similarity between the eukaryotic TATA box and the prokaryotic −10 box. The major difference between the two is position with respect to the transcription start site: −25 to −30 versus −10. (TATA boxes in yeast [*Saccharomyces cerevisiae*] have a more variable location, from 30 to more than 300 bp upstream of their transcription start sites.)

As usual with consensus sequences, exceptions to the rule exist. Indeed, in this case they are plentiful. Sometimes G's and C's creep in, as in the TATA box of the rabbit β-globin gene, which starts with the sequence CATA. Frequently, no recognizable TATA box is evident at all. Such TATA-less promoters tend to be found in two classes of genes: (1) The first class comprises the **housekeeping genes** that are constitutively active in virtually all cells because they control common biochemical pathways, such as nucleotide synthesis, needed to sustain cellular life. Thus, we find TATA-less promoters in the cellular genes for adenine deaminase, thymidylate synthetase, and dihydrofolate reductase, all of which encode enzymes necessary for making nucleotides, and in the SV40 region encoding the viral late proteins. These genes sometimes have GC boxes that appear to compensate for the lack of a TATA box (Chapter 11). In *Drosophila,* only about 30% of class II promoters have recognizable TATA boxes, but many TATA-less promoters have DPEs that play the same role as a TATA box. (2) The second class of genes with TATA-less promoters are developmentally regulated genes such as the homeotic genes that control development of the fruit fly or genes that are active during development of the immune system in mammals. We will examine one such gene (the mouse terminal deoxynucleotidyltransferase [TdT] gene) later in this chapter. In general, **specialized genes** (sometimes called *luxury genes*), which encode proteins made only in certain types of cells (e.g., keratin in skin cells and hemoglobin in red blood cells), do have TATA boxes.

What is the function of the TATA box? That seems to depend on the gene. The first experiments to probe this question involved deleting the TATA box and then assaying the deleted DNA for promoter activity by transcription in vitro.

In 1981, Christophe Benoist and Pierre Chambon performed a deletion mutagenesis study of the SV40 early promoter. The assays they used for promoter activity were primer extension and S1 mapping. These techniques, described in Chapter 5, produce labeled DNA fragments whose lengths tell us where transcription starts and whose abundance tells us how active the promoter is. As Figure 10.19a shows, the P1A, AS, HS0, HS3, and HS4 mutants, which Benoist and Chambon had created by deleting progressively more of the DNA downstream of the TATA box, including the initiation site, simply shortened the S1 signal by an amount equal to the number of base pairs removed by the deletion. This result is consistent with a downstream shift in the transcription start site caused by the deletion. Such a shift is just what we would predict if the TATA box positions transcription initiation approximately 25 to 30 bp downstream of the last base of the TATA box. If this is so, what should be the consequences of deleting the TATA box altogether? The H2 deletion extends the H4 deletion through the TATA box and therefore provides the answer to our question: Lane 8 of Figure 10.19b shows that removing the TATA box caused transcription to initiate at a wide variety of sites, while not decreasing the efficiency of transcription. If anything, the darkness of the S1 signals suggests an increase in transcription. Thus, it appears that the TATA box is involved in positioning the start of transcription.

In further experiments, Benoist and Chambon reinforced this conclusion by systematically deleting DNA between the TATA box and the initiation site of the SV40 early gene and locating the start of transcription in the resulting shortened DNAs by S1 mapping. Transcription of the wild-type gene begins at three different guanosines, clustered 27–34 bp downstream of the first T of the TATA box. As Benoist and Chambon removed more and more of the DNA between the TATA box and these initiation sites, they noticed that transcription no longer initiated at these sites. Instead, transcription started at other bases, usually purines, that

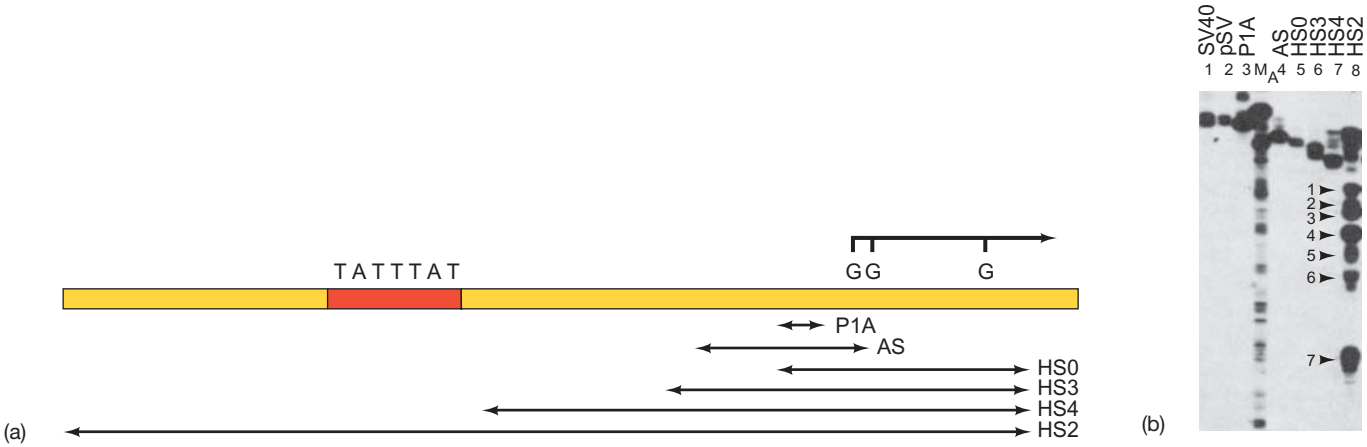


Figure 10.19 Effects of deletions in the SV40 early promoter. **(a)** Map of the deletions. The names of the mutants are given at the right of each arrow. The arrows indicate the extent of each deletion. The positions of the TATA box (TATTTAT, red) and the three transcription start sites (all G's) are given at top. **(b)** Locating the transcription start sites in the mutants. Benoist and Chambon transfected cells with either SV40 DNA, or a plasmid containing the wild-type SV40 early region (pSV1), or a derivative of pSV1 containing one of the mutated SV40 early promoters described in panel (a). They located the initiation site (or sites) by S1 mapping. The names of the mutants being tested are given at the top of each lane. The lane denoted M_A contained size markers. The numbers to the left of the bands in the HS2 lane denote novel transcription start sites not detected with the wild-type promoter or with any of the other mutants in this experiment. The heterogeneity in the transcription initiation sites was apparently due to the lack of a TATA box in this mutant. (*Source:* (*b*) Benoist C. and P. Chambon, In vivo sequence requirements of the SV40 early promoter region. *Nature* 290 (26 Mar 1981) p. 306, f. 3.)

were about 30 bp downstream of the first T of the TATA box. In other words, the distance between the TATA box and the transcription initiation sites remained constant, with little regard to the exact sequence at these initiation sites.

In this example, the TATA box appears to be important for locating the start of transcription, but not for regulating the efficiency of transcription. However, in some other promoters, removal of the TATA box impairs promoter function to such an extent that transcription, even from aberrant start sites, cannot be detected.

Steven McKnight and Robert Kingsbury provided an example with their studies of the herpes virus thymidine kinase (tk) promoter. They performed **linker scanning mutagenesis,** in which they systematically substituted a synthetic 10-bp linker for 10-bp sequences throughout the tk promoter. One of the results of this analysis was that mutations within the TATA box destroyed promoter activity (Figure 10.20). In the mutant with the lowest promoter activity (LS –29/–18), the normal sequence in the region of the TATA box had been changed from GCATATTA to CCGGATCC.

Thus, some class II promoters require the TATA box for function, but others need it only to position the transcription start site. And, as we have seen, some class II promoters, most notably the promoters of housekeeping genes, have no TATA box at all, and they still function quite well. How do we account for these differences? As we will see in Chapters 11 and 12, promoter activity depends on assembling a collection of transcription factors and RNA polymerase called a preinitiation complex. This complex forms at the transcription start site and launches the transcription process. In class II promoters, the TATA box serves as the site where this assembly of protein factors begins. The first protein to bind is TFIID, including the TATA-box-binding protein (TBP), which then attracts the other factors. But what about promoters that lack TATA boxes? These still require TBP, but because TBP has no TATA box to which it can bind, it depends on other proteins, which bind to other promoter elements, to hold it in place.

Initiators, Downstream Promoter Elements, and TFIIB Recognition Elements Some class II promoters have conserved sequences around their transcription start sites that are required for optimal transcription. These are called **initiators,** and mammalian initiators have the consensus sequence PyPyAN(T/A)PyPy, where Py stands for either pyrimidine (C or T), N stands for any base, and the underlined A is the transcription start point. *Drosophila* initiators have the consensus sequence TCA(G/T)T(T/C). The classic example of an initiator comes from the adenovirus major late promoter. This initiator, together with the TATA box, constitutes a core promoter that can drive transcription of any gene placed downstream of it, though at a very low level. This promoter is also susceptible to stimulation by upstream elements or enhancers connected to it.

Another example of a gene with an important initiator is the mammalian terminal deoxynucleotidyltransferase (TdT) gene, which is activated during development of B and T lymphocytes. Stephen Smale and David Baltimore studied the mouse TdT promoter and found that it contains no TATA box and no apparent upstream promoter elements, but it does contain an initiator. This initiator is sufficient to drive basal-level transcription of the gene from a single start

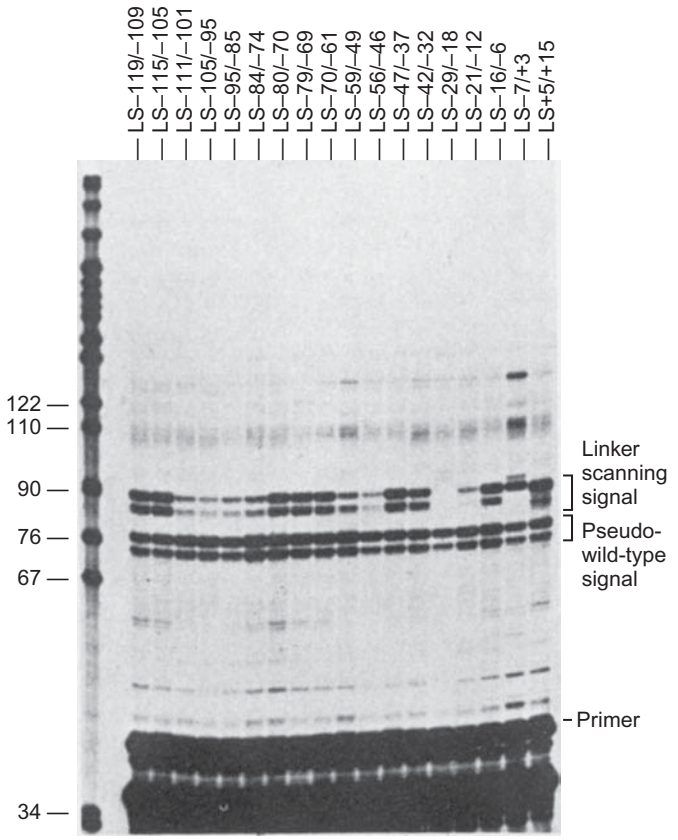


Figure 10.20 Effects of linker scanning mutations in the herpes virus tk promoter. McKnight and Kingsbury made linker scanning mutations throughout the tk promoter, then injected the mutated DNAs into frog oocytes, along with a pseudo-wild-type DNA (mutated at the +21 to +31 position). Transcription from this pseudo-wild-type promoter was just as active as that from the wild-type promoter, so this DNA served as an internal control. The investigators assayed for transcription from the test plasmid and from the control plasmid by primer extension analysis. Transcription from the control plasmid remained relatively constant, as expected, but transcription from the test plasmid varied considerably depending on the locus of the mutations. (*Source:* Adapted from McKnight, S.L. and R. Kingsbury, Transcriptional control signals of a eukaryotic protein-coding gene. *Science* 217 (23 July 1982) p. 322, f. 5.)

site located within the initiator sequence. Smale and Baltimore also found that a TATA box or the GC boxes from the SV40 promoter could greatly stimulate transcription starting at the initiator. Thus, this initiator alone constitutes a very simple, but functional, promoter whose efficiency can be enhanced by other promoter elements.

Downstream promoter elements are very common in *Drosophila*. In fact, in 2000 Alan Kutach and James Kadonaga reported the surprising discovery that DPEs are just as common in *Drosophila* as TATA boxes. These DPEs are found about 30 bp downstream of the transcription initiation site and include the consensus sequence G(A/T)CG. They can compensate for the loss of the TATA box from a promoter. Indeed, many naturally TATA-less promoters in *Drosophila* contain DPEs, which accounts for the abundance of DPEs in this organism. It is common to find a DPE coupled with an Inr in TATA-less *Drosophila* promoters. The similarity between the TATA box and the DPE extends to their ability to bind to a key general transcription factor known as TFIID (Chapter 11).

Another important general transcription factor is TFIIB, which binds to the promoter along with TFIID, RNA polymerase II, and other factors, to form a preinitiation complex that is competent to begin transcription. Some promoters have a DNA element just upstream of the TATA box that helps TFIIB to bind to the DNA. These are called TFIIB recognition elements (BREs).

SUMMARY Class II promoters may consist of a core promoter immediately surrounding the transcription start site, and a proximal promoter further upstream. The core promoter may contain up to six conserved elements: the TFIIB recognition element (BRE), the TATA box, the initiator (Inr), the downstream core element (DCE), the motif ten element (MTE), and the downstream promoter element (DPE). At least one of these elements is missing in most promoters. In fact, TATA-less promoters tend to have DPEs, at least in *Drosophila*. Promoters for highly expressed specialized genes tend to have TATA boxes, but promoters for housekeeping genes tend to lack them.

Proximal Promoter Elements McKnight and Kingsbury's linker scanning analysis of the herpes virus tk gene revealed other important promoter elements upstream of the TATA box. Figure 10.20 shows that mutations in the −47 to −61 and in the −80 to −105 regions caused significant loss of promoter activity. The nontemplate strands of these regions contain the sequences GGGCGG and CCGCCC, respectively. These are so-called **GC boxes,** which are found in a variety of promoters, usually upstream of the TATA box. Notice that the two GC boxes are in opposite orientations in their two locations in the herpes virus tk promoter.

Chambon and colleagues also found GC boxes in the SV40 early promoter, and not just two copies, but six. Furthermore, mutations in these elements significantly decreased promoter activity. For example, loss of one GC box decreased transcription to 66% of the wild-type level, and loss of a second GC box decreased transcription all the way down to 13% of the control level. We will see in Chapter 12 that a specific transcription factor called Sp1 binds to the GC boxes and stimulates transcription. Later in this chapter we will discuss DNA elements called enhancers that stimulate transcription, but differ from promoters in two important respects: They are position- and orientation-independent. The GC boxes are orientation-independent; they can be flipped 180 degrees and they still function (as occurs naturally in the herpes virus tk promoter). But

the GC boxes do not have the position independence of classical enhancers, which can be moved as much as several kilobases away from a promoter, even downstream of a gene's coding region, and still function. If the GC boxes are moved more than a few dozen base pairs away from their own TATA box, they lose the ability to stimulate transcription. Thus, it is probably more proper to consider the GC boxes, at least in these two genes, as proximal promoter elements, rather than enhancers. On the other hand, the distinction is subtle and perhaps borders on semantic.

Another upstream element found in a wide variety of class II promoters is the so-called **CCAAT box** (pronounced "cat box"). In fact, the herpes virus tk promoter has a CCAAT box; the linker scanning study we have discussed failed to detect any loss of activity when this CCAAT box was mutated, but other investigations have clearly shown the importance of the CCAAT box in this and in many other promoters. Just as the GC box has its own transcription factor, so the CCAAT box must bind a transcription factor (the **CCAAT-binding transcription factor [CTF],** among others) to exert its stimulatory influence.

SUMMARY Proximal promoter elements are usually found upstream of class II core promoters. They differ from the core promoter in that they bind to relatively gene-specific transcription factors. For example, GC boxes bind the transcription factor Sp1, while CCAAT boxes bind CTF. The proximal promoter elements, unlike the core promoter, can be orientation-independent, but they are relatively position-dependent, unlike classical enhancers.

Class I Promoters

What about the promoter recognized by RNA polymerase I? We can refer to this promoter in the singular because almost all species have only one kind of gene recognized by polymerase I: the rRNA precursor gene. The one known exception is the trypanosome, in which polymerase I transcribes two protein-encoding genes, in addition to the rRNA precursor gene. It is true that the rRNA precursor gene is present in hundreds of copies in each cell, but each copy is virtually the same as the others, and they all have the same promoter sequence. However, this sequence is quite variable from one species to another—more variable than those of the promoters recognized by polymerase II, which tend to have conserved elements, such as TATA boxes, in common.

Robert Tjian and colleagues used linker scanning mutagenesis to identify the important regions of the human rRNA promoter. Figure 10.21 shows the results of this analysis: The promoter has two critical regions in which mutations cause a great reduction in promoter strength. One of these, the **core element,** also known at the initiator (rINR), is located at the start of transcription, between positions −45 and +20. The other is the **upstream promoter element (UPE),** located between positions −156 and −107.

The presence of two promoter elements raises the question of the importance of the spacing between them. In this case, spacing is very important. Tjian and colleagues deleted or added DNA fragments of various lengths between the UPE and the core element of the human rRNA promoter. When they removed only 16 bp between the two promoter elements, the promoter

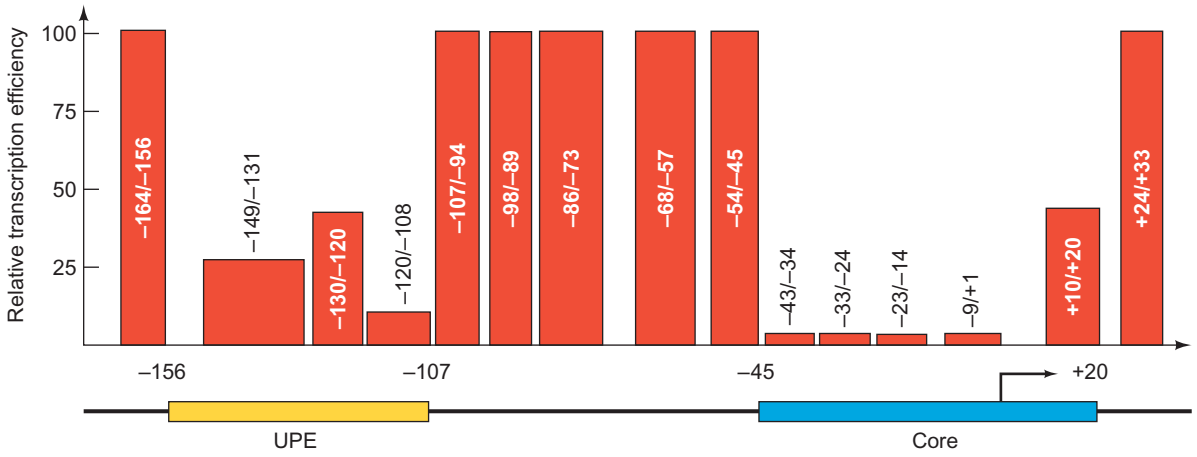


Figure 10.21 Two rRNA promoter elements. Tjian and colleagues used linker scanning to mutate short stretches of DNA throughout the 5′-flanking region of the human rRNA gene. They then tested these mutated DNAs for promoter activity using an in vitro transcription assay. The bar graph illustrates the results, which show that the promoter has two important regions: labeled UPE (upstream promoter element) and Core. The UPE is necessary for optimal transcription, but basal transcription is possible in its absence. On the other hand, the core element is absolutely required for any transcription to occur. (*Source:* Adapted from Learned, R.M., T.K. Learned, M.M. Haltiner, and R.T. Tjian, Human rRNA transcription is modulated by the coordinated binding of two factors to an upstream control element. *Cell* 45:848, 1986.)

strength dropped to 40% of wild-type; by the time they had deleted 44 bp, the promoter strength was only 10%. On the other hand, they could add 28 bp between the elements without affecting the promoter, but adding 49 bp reduced promoter strength by 70%. Thus, the promoter efficiency is more sensitive to deletions than to insertions between the two promoter elements.

SUMMARY Class I promoters are not well conserved in sequence from one species to another, but the general architecture of the promoter is well conserved. It consists of two elements, a core element surrounding the transcription start site, and an upstream promoter element (UPE) about 100 bp farther upstream. The spacing between these two elements is important.

Class III Promoters

As we have seen, RNA polymerase III transcribes a variety of genes that encode small RNAs. These include (1) the "classical" class III genes, including the 5S rRNA and tRNA genes, and the adenovirus VA RNA genes; and (2) some relatively recently discovered class III genes, including the U6 snRNA gene, the 7SL RNA gene, the 7SK RNA gene, and the Epstein–Barr virus EBER2 gene. The latter, "nonclassical" class III genes have promoters that resemble those found in class II genes. By contrast, the "classical" class III genes have promoters located entirely within the genes themselves.

Class III Genes with Internal Promoters Donald Brown and his colleagues performed the first analysis of a class III promoter, on the gene for the *Xenopus borealis* 5S rRNA. The results they obtained were astonishing. Whereas the promoters recognized by polymerases I and II, as well as by bacterial polymerases, are located mostly in the 5′-flanking region of the gene, the 5S rRNA promoter is located *within* the gene it controls.

The experiments that led to this conclusion worked as follows: First, to identify the 5′-end of the promoter, Brown and colleagues prepared a number of mutant 5S rRNA genes that were missing more and more of their 5′-end and observed the effects of the mutations on transcription in vitro. They scored transcription as correct by measuring the size of the transcript by gel electrophoresis. An RNA of approximately 120 bases (the size of 5S rRNA) was deemed an accurate transcript, even if it did not have the same sequence as real 5S rRNA. They had to allow for incorrect sequence in the transcript because they changed the internal sequence of the gene to disrupt the promoter.

The surprising result (Figure 10.22) was that the entire 5′-flanking region of the gene could be removed without affecting transcription very much. Furthermore, big chunks of the 5′-end of the gene itself could be removed, and a transcript of about 120 nt would still be made. However, deletions beyond about position +50 destroyed promoter function.

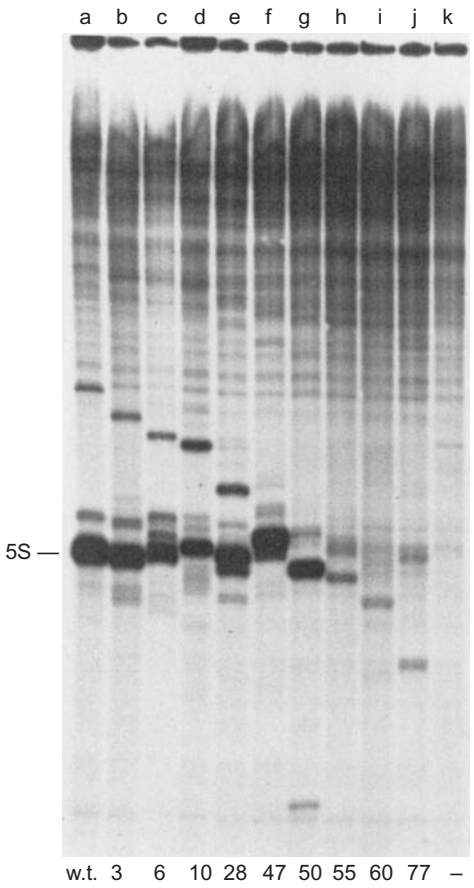


Figure 10.22 Effect of 5′-deletions on 5S rRNA gene transcription. Brown and colleagues prepared a series of deleted *Xenopus borealis* 5S rRNA genes with progressively more DNA deleted from the 5′-end of the gene itself. Then they transcribed these deleted genes in vitro in the presence of labeled substrate and electrophoresed the labeled products. DNA templates: lane a, undeleted positive control; lanes b–j, deleted genes with the position of the remaining 5′-end nucleotide denoted at bottom (e.g., lane b contained the product of a 5S rRNA gene whose 5′-end is at position +3 relative to the wild-type gene); lane k, negative control (pBR322 DNA with no 5S rRNA gene). Strong synthesis of a 5S-size RNA took place with all templates through lane g, in which deletion up to position +50 had occurred. With further deletion into the gene, this synthesis ceased. Lanes h–k also contained a band in this general area, but it is an artifact unrelated to 5S rRNA gene transcription. (*Source:* Sakonju, S., D.F. Bogenhagen, and D.D. Brown. 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 (Jan 1980) p. 17, f. 4.)

Using a similar approach, Brown and colleagues identified a sensitive region between bases 50 and 83 of the transcribed sequence that could not be encroached on without destroying promoter function. These are the apparent outer

boundaries of the internal promoter of the *Xenopus* 5S rRNA gene. Other experiments showed that it is possible to *add* chunks of DNA outside this region without harming the promoter. Roeder and colleagues later performed systematic mutagenesis of bases throughout the promoter region and identified three regions that could not be changed without greatly diminishing promoter function. These sensitive regions are called *box A*, the *intermediate element*, and *box C*. (No box B occurs because a box B had already been discovered in other class III genes, and it had no counterpart in the 5S rRNA promoter.) Figure 10.23a summarizes the results of these experiments on the 5S rRNA promoter. Similar experiments on the other two classical class III genes, the tRNA and VA RNA genes, showed that their promoters contain a *box A* and a *box B* (Figure 10.23b). The sequence of the box A is similar to that of the box A of the 5S rRNA gene. Furthermore, the space in between the two blocks can be altered somewhat without destroying promoter function. Such alteration does have limits, however; if one inserts too much DNA between the two promoter boxes, efficiency of transcription suffers.

Thus, we see that there are several kinds of class III promoters. The 5S rRNA genes are in a group by themselves, called *type I* (Figure 10.23a). Do not confuse this with "class I;" we are discussing only class III promoters here. The second group, *type II*, contains most class III promoters, which look like the tRNA and VA RNA promoters in Figure 10.23b. The third group, *type III*, contains the nonclassical promoters with control elements restricted to the 5′-flanking region of the gene. These, promoters are typified by the human 7SK RNA promoter and the human U6 RNA promoter (Figure 10.23c). By the way, the U6 RNA is a member of a group of small nuclear RNAs (snRNAs) that are key players in mRNA splicing, which we will discuss in Chapter 14. Finally, there are promoters that appear to be hybrids of types II and III, such as the human 7SL promoter. These have both internal and external elements that are important for promoter activity.

SUMMARY RNA polymerase III transcribes a set of short genes. The classical class III genes (types I and II) have promoters that lie wholly within the genes. The internal promoter of the type I class III gene (the 5S rRNA gene) is split into three regions: box A, a short intermediate element, and box C. The internal promoters of the type II genes (e.g., the tRNA genes) are split into two parts: box A and box B. The promoters of the nonclassical (type III) class III genes resemble those of class II genes.

Class III Genes with Class II-like Promoters After Brown and other investigators established the novel idea of internal promoters for class III genes, it was generally assumed that all class III genes worked this way. However, by the mid-1980s some exceptions were discovered. The **7SL RNA** is part of the signal recognition particle that recognizes a signal sequence in certain mRNAs and targets their translation to membranes such as the endoplasmic reticulum. In 1985, Elisabetta Ullu and Alan Weiner conducted in vitro transcription studies on wild-type and mutant 7SL RNA genes that showed that the 5′-flanking region was required for high-level transcription. Without this DNA region, transcription efficiency dropped by 50–100-fold. Ullu and Weiner concluded that the most important DNA element for transcription of this gene lies upstream of the gene. Nevertheless, the fact that transcription still occurred in mutant genes lacking the 5′-flanking region implies that these genes also contain a weak internal promoter. These data help explain why the hundreds

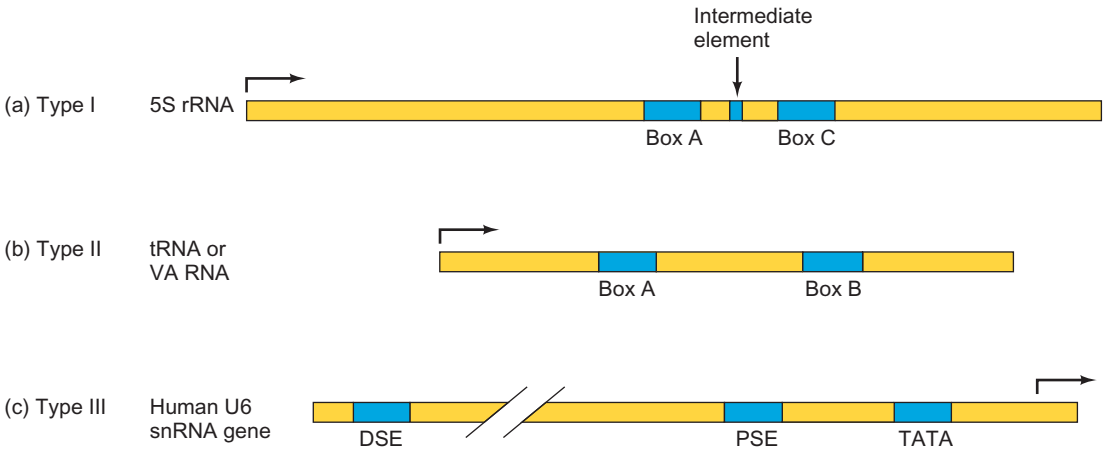


Figure 10.23 Promoters of some class III genes. The promoters of the 5S, tRNA and U6 RNA genes are depicted as groups of blue boxes within the genes they control. DSE and PSE are distal and proximal sequence elements, respectively.

of 7SL RNA pseudogenes (nonfunctional copies of the 7SL gene) in the human genome, as well as the related *Alu* sequences (remnants of transposons, Chapter 23), are relatively poorly transcribed in vivo: They lack the upstream element required for high-level transcription.

Marialuisa Melli and colleagues noticed that the 7SK RNA gene does not have internal sequences that resemble the classic class III promoter. On the other hand, the 7SK RNA gene does have a 5′-flanking region homologous to that of the 7SL RNA gene. On the basis of these observations, they proposed that this gene has a completely external promoter. To prove the point, they made successive deletions in the 5′-flanking region of the gene and tested them for ability to support transcription in vitro. Figure 10.24 shows that deletions up to position −37 still allowed production of high levels of 7SK RNA, but deletions downstream of this point were not tolerated. On the other hand, the coding region was not needed for transcription: In vitro transcription analysis of another batch of deletion mutants, this time with deletions within the coding region, showed that transcription still occurred, even when the whole coding region was removed. Thus, this gene lacks an internal promoter.

What is the nature of the promoter located in the region encompassing the 37 bp upstream of the start site? Interestingly enough, a TATA box resides in this region, and changing three of its bases (TAT→GCG) reduced transcription by 97%. Thus the TATA box is required for good promoter function. All this may make you wonder whether polymerase II, not polymerase III, really transcribes this gene after all. If that were the case, low concentrations of α-amanitin should inhibit transcription, but it takes high concentrations of this toxin to block 7SK RNA synthesis. In fact, the profile of inhibition of 7SK RNA synthesis by α-amanitin is exactly what we would expect if polymerase III, not polymerase II, is involved. By the way, the **7SK RNA** plays a role in controlling the phosphorylation of one serine (serine 2) in the repeating heptad of the CTD of Rpb1 of RNA polymerase II. We will see in Chapter 11 that this phosphorylation is required for the transition from transcription initiation to elongation.

Now we know that the other nonclassical class III genes, including the U6 RNA gene and the EBER2 gene, behave the same way. They are transcribed by polymerase III, but they have polymerase II-like promoters. In Chapter 11 we will see that this is not as strange as it seems at first because the TATA-binding protein (TBP) is involved in class III (and class I) transcription, in addition to its well-known role in class II gene transcription.

The small nuclear RNA (snRNA) genes present a fascinating comparison of class II and class III nonclassical promoters. In Chapter 14 we will learn that many eukaryotic mRNAs are synthesized as over-long precursors that need to have internal sections (introns) removed in a process called splicing. This pre-mRNA splicing requires several small nuclear RNAs (snRNAs). Most of these, including U1 and U2 snRNAs, are made by RNA polymerase II. But their promoters do not look like typical class II promoters. Instead, in humans, each promoter contains two elements (Figure 10.25a): a **proximal sequence element** (**PSE**), which is essential, and a **distal sequence element** (**DSE**), which confers greater efficiency.

One of the snRNAs, U6 snRNA, is made by RNA polymerase III. As usual with nonclassical class III promoters, the human U6 snRNA promoter (Figure 10.25b), with its TATA

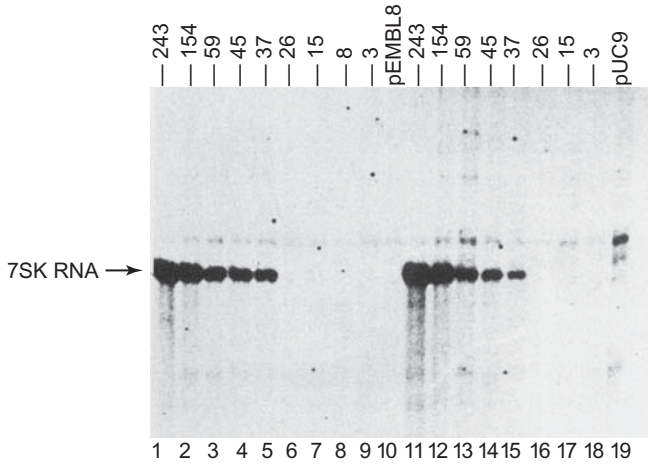


Figure 10.24 Effects of 5′-deletion mutations on the 7SK RNA promoter. Melli and colleagues performed deletions in the 5′-flanking region of the human 7SK RNA gene and transcribed the mutated genes in vitro. Then they electrophoresed the products to determine if 7SK RNA was still synthesized. The negative numbers at the top of each lane give the number of base pairs of the 5′-flanking region still remaining in the deleted gene used in that reaction. For example, the template used in lane 9 retained only 3 bp of the 5′-flanking region—up to position −3. Lanes 1–10 contained deleted genes cloned into the vector pEMBL8; lanes 11–19 contained genes cloned into pUC9. The cloning vectors themselves were transcribed in lanes 10 and 19. A comparison of lanes 5 and 6 (or of lanes 15 and 16) shows an abrupt drop in promoter activity when the bases between position −37 and −26 were removed. This suggests that an important promoter element lies in this 11-bp region. (*Source:* Murphy, S., C. DiLiegro, and M. Melli, The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51 (9) (1987) p. 82, f. 1b.)

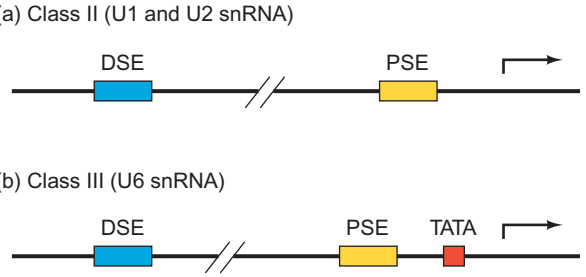


Figure 10.25 Structures of class II and III nonclassical promoters. **(a)** Class II: The U1 and U2 snRNA promoters contain an essential PSE near the transcription start site and a supplementary DSE further upstream. **(b)** Class III: The U6 snRNA promoter contains a TATA box in addition to the PSE and DSE.

box, looks more like a class II promoter. Paradoxically, removal of that TATA box converts the promoter from class III to class II. Similarly, adding a TATA box to a U1 or U2 snRNA promoter converts it from class II to class III. One might have predicted just the opposite. By contrast, in *Drosophila* and in sea urchins, some snRNA genes have TATA boxes and others do not, but other sequence elements, not the TATA boxes, determine whether the promoters are class II or class III.

SUMMARY At least one class III gene, the 7SL RNA gene, contains a weak internal promoter, as well as a sequence in the 5′-flanking region of the gene that is required for high-level transcription. Other nonclassical class III genes (e.g., 7SK, and U6 RNA genes) lack internal promoters altogether, and contain promoters that strongly resemble class II promoters in that they lie in the 5′-flanking region and contain TATA boxes. The U1 and U6 snRNA genes have nonclassical class II and III promoters, respectively. The U1 snRNA promoter has an essential proximal sequence element (PSE), and a distal sequence element (DSE) and is transcribed by polymerase II. The U6 snRNA promoter has a PSE, a DSE, and a TATA box, and is transcribed by polymerase III.

10.3 Enhancers and Silencers

Many eukaryotic genes, especially class II genes, are associated with *cis*-acting DNA elements that are not strictly part of the promoter, yet strongly influence transcription. As we learned in Chapter 9, **enhancers** are elements that stimulate transcription. **Silencers,** by contrast, depress transcription. We will discuss these elements briefly here and expand on their modes of action in Chapters 12 and 13.

Enhancers

Chambon and colleagues discovered the first enhancer in the 5′-flanking region of the SV40 early gene. This DNA region had been noticed before because it contains a conspicuous duplication of a 72-bp sequence, called the *72-bp repeat* (Figure 10.26). When Benoist and Chambon made deletion mutations in this region, they observed profoundly depressed transcription in vivo. This behavior suggested that the 72-bp repeats constituted another upstream promoter element. However, Paul Berg and his colleagues discovered that the 72-bp repeats still stimulated transcription even if they were inverted or moved all the way around to the opposite side of the circular SV40 genome, over 2 kb away from the promoter. The latter behavior, at least, is very un-promoter-like. Thus, such orientation- and position-independent DNA elements are called enhancers to distinguish them from promoter elements.

How do enhancers stimulate transcription? We will see in Chapter 12 that enhancers act through proteins that bind to them. These have several names: **transcription factors, enhancer-binding proteins,** or **activators.** These proteins appear to stimulate transcription by interacting with other proteins called **general transcription factors** at the promoter. This interaction promotes formation of a preinitiation complex, which is necessary for transcription. Thus, enhancers usually allow a gene to be induced (or sometimes repressed) by activators. We will discuss these interactions in much greater detail in Chapters 11 and 12 and we will see that activators frequently require help from other molecules (e.g., hormones and coactivator proteins) to exert their effects.

We frequently find enhancers upstream of the promoters they control, but this is by no means an absolute rule. In fact, as early as 1983 Susumo Tonegawa and his colleagues found an example of an enhancer *within* a gene. These investigators were studying a gene that encodes the larger subunit of a particular mouse antibody, or immunoglobulin, called γ_{2b}. They introduced this gene into mouse plasmacytoma cells that normally expressed antibody genes, but not this particular gene. To detect efficiency of expression of the transfected cells, they added a labeled amino acid to tag newly made proteins, then **immunoprecipitated** the labeled γ_{2b} protein (Chapter 5) with an antibody directed against γ_{2b}. Then they electrophoresed the immunoprecipitated proteins and detected them by autoradiography. The suspected enhancer lay in one of the gene's introns, a region within the gene that is transcribed, but is subsequently cut out of the transcript by a process called splicing (Chapter 14). Tonegawa and colleagues began by deleting two chunks of DNA from this suspected enhancer region, as shown in Figure 10.27a. Then they assayed for expression of the γ_{2b} gene in cells transfected by this mutated DNA. Figure 10.27b shows the results: The deletions within the intron, though they should have no effect on the protein product because they are in a noncoding region of

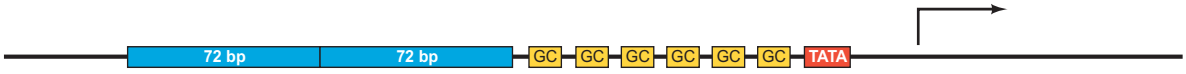


Figure 10.26 Structure of the SV40 virus early control region. As usual, an arrow with a right-angle bend denotes the transcription initiation site, although this is actually a cluster of three sites, as we saw in Figure 10.19. Upstream of the start site we have, in right-to-left order, the TATA box (red), six GC boxes (yellow), and the enhancer (72-bp repeats, blue).

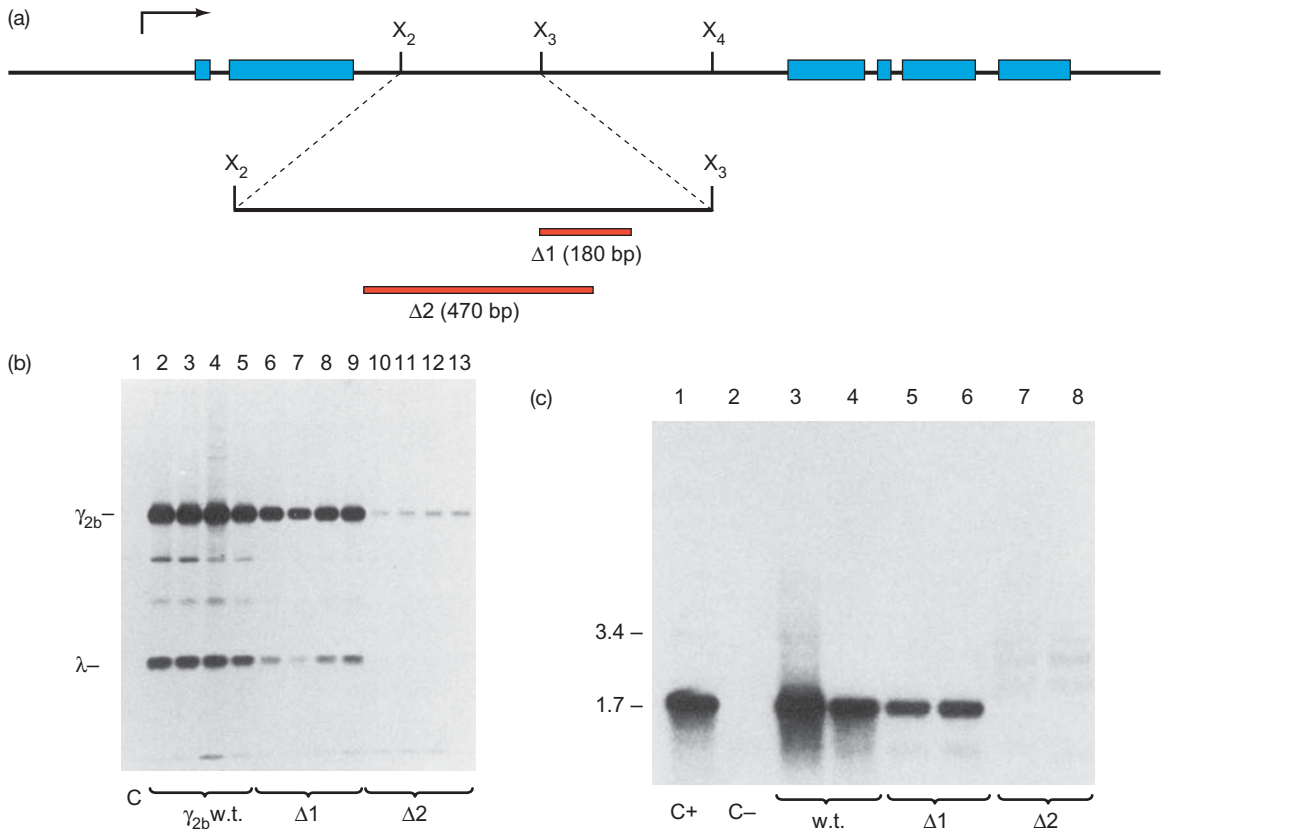


Figure 10.27 Effects of deletions in the immunoglobulin γ_{2b} H-chain enhancer. (a) Map of the cloned γ_{2b} gene. The blue boxes represent the exons of the gene, the parts that are included in the mRNA that comes from this gene. The lines in between boxes are introns, regions of the gene that are transcribed, but then cut out of the mRNA precursor as it is processed to the mature mRNA. X_2, X_3, and X_4 represent cutting sites for the restriction enzyme *Xba*I. Tonegawa and colleagues suspected an enhancer lay in the X_2–X_3 region, so they made deletions Δ1 and Δ2 as indicated by the red boxes. **(b)** Assay of expression of the γ_{2b} gene at the protein level. Tonegawa and colleagues transfected plasmacytoma cells with the wild-type gene (lanes 2–5), the gene with deletion Δ1 (lanes 6–9), or the gene with deletion Δ2 (lanes 10–13). Lane 1 was a control with untransfected plasmacytoma cells. After transfecting the cells, these investigators added a radioactive amino acid to label any newly made protein, then extracted the protein, immunoprecipitated the γ_{2b} protein, electrophoresed the precipitated protein, and detected the radioactive protein by fluorography (a modified version of autoradiography in which a compound called a fluor is added to the electrophoresis gel). Radioactive emissions excite this fluor to give off photons that are detected by x-ray film. The Δ1 deletion produced only a slight reduction in expression of the gene, but the Δ2 deletion gave a profound reduction. **(c)** Assay of transcription of the γ_{2b} gene. Tonegawa and colleagues electrophoresed and Northern blotted RNA from the following cells: lane 1 (positive control), untransfected plasmacytoma cells (MOPC 141) that expressed the γ_{2b} gene; lane 2 (negative control), untransfected plasmacytoma cells (J558L) that did not express the γ_{2b} gene; lanes 3 and 4, J558L cells transfected with the wild-type γ_{2b} gene; lanes 5 and 6, J558L cells transfected with the gene with the Δ1 deletion; lanes 7 and 8, J558L cells transfected with the gene with the Δ2 deletion. The Δ1 deletion decreased transcription somewhat, but the Δ2 deletion abolished transcription. (*Source:* (*b–c*) Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa, A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33 (July 1983) p. 719, f. 2&3.)

the gene, caused a decrease in the amount of gene product made. This was especially pronounced in the case of the larger deletion (Δ2).

Is this effect due to decreased transcription, or some other cause? Tonegawa's group answered this question by performing Northern blots (Chapter 5) with RNA from cells transfected with normal and deleted γ_{2b} genes. These blots, shown in Figure 10.27c, again demonstrated a profound loss of function when the suspected enhancer was deleted. But is this really an enhancer? If so, one should be able to move it or invert it and it should retain its activity. Tonegawa and colleagues did this by first inverting the X_2–X_3 fragment, which contained the enhancer, as shown in position/orientation B of Figure 10.28a. Figure 10.28b shows that the enhancer still functioned. Next, they took fragment X_2–X_3 out of the intron and placed it upstream of the promoter (position/orientation C). It still worked. Then they inverted it in its new location (position/orientation D). Still it functioned. Thus, some region within the X_2–X_3 fragment behaved as an enhancer: It stimulated transcription from a nearby promoter, and it was position- and orientation-independent.

Finally, these workers compared the expression of this gene when it was transfected into two different types of mouse cells: plasmacytoma cells as before, and fibroblasts.

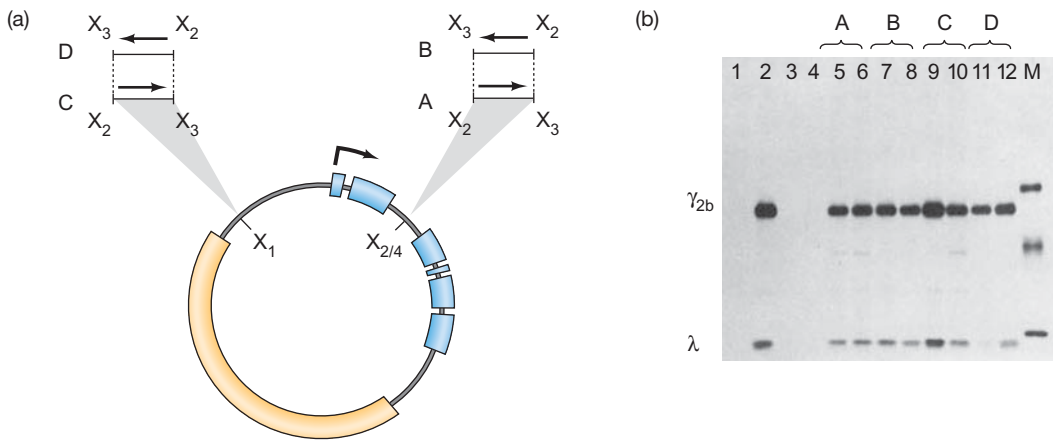


Figure 10.28 The enhancing element in the γ_{2b} gene is orientation- and position-independent. (a) Outline of the mutant plasmids. Tonegawa and colleagues removed the X_2–X_3 region of the parent plasmid containing the γ_{2b} gene (see Figure 10.27a). This deleted the enhancer. Then they reinserted the X_2–X_3 fragment (with the enhancer) in four different ways: plasmids A and B, the fragment was inserted back into the intron in its usual location in the forward (normal) orientation (A), or in the backward orientation (B); plasmids C and D, the fragment was inserted into another *Xba*I site (X_1) hundreds of base pairs upstream of the gene in the forward orientation (C), or in the backward orientation (D). **(b)** Experimental results. Tonegawa and colleagues tested all four plasmids from (a), as well as the parent, for efficiency of expression as in Figure 10.27b. All functioned equally well. Lane 1, untransfected J558L cells lacking the γ_{2b} gene. Lanes 2–12, J558L cells transfected with the following plasmids: lane 2, the parent plasmid with no deletions; lanes 3 and 4, the parent plasmid with the X_2–X_3 fragment deleted; lanes 5 and 6, plasmid A; lanes 7 and 8, plasmid B; lanes 9 and 10, plasmid C; lanes 11 and 12, plasmid D. Lane M contained protein size markers. (*Source:* (a) Adapted from Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa, A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33 (July 1983) p. 721, f. 5.)

Expression was much more active in plasmacytoma cells. This is also consistent with enhancer behavior because fibroblasts do not make antibodies and therefore should not contain enhancer-binding proteins capable of activating the enhancer of an antibody gene. Thus, the antibody gene should not be expressed actively in such cells.

The finding that a gene is much more active in one cell type than in another leads to an extremely important point: All cells contain the same genes, but different cell types differ greatly from one another: A nerve cell, for example, is much different from a liver cell, in shape and function. What makes these cells differ so much? The proteins in the cells. And, as we have learned, the suite of proteins in each cell type is determined by the genes that are active in those cells. And what activates those genes? We now see that the activators are transcription factors that bind to enhancers. Thus, different cell types express different activators that turn on different genes that produce different proteins. We will expand on this vital theme in several chapters to follow.

Silencers

Enhancers are not the only DNA elements that can act at a distance to modulate transcription. Silencers also do this, but—as their name implies—they inhibit rather than stimulate transcription. The mating system *(MAT)* of yeast provides a good example. Yeast chromosome III contains three loci of very similar sequence: *MAT, HML,* and *HMR*. Though *MAT* is expressed, the other two loci are not, and silencers located at least 1 kb away seem to be responsible for this genetic inactivity. We know that something besides the inactive genes themselves is at fault, because active yeast genes can be substituted for *HML* or *HMR* and the transplanted genes become inactive. Thus, they seem to be responding to an external negative influence: a silencer. How do silencers work? The available data indicate that they cause the chromatin to coil up into a condensed, inaccessible, and therefore inactive form, thereby preventing transcription of neighboring genes. We will examine this process in more detail in Chapter 13.

Sometimes the same DNA element can have both enhancer and silencer activity, depending on the protein bound to it. For example, the thyroid hormone response element acts as a silencer when the thyroid hormone receptor binds to it without its ligand, thyroid hormone. But it acts as an enhancer when the thyroid hormone receptor binds along with thyroid hormone. We will revisit this concept in Chapter 12.

SUMMARY Enhancers and silencers are position- and orientation-independent DNA elements that stimulate or depress, respectively, the transcription of associated genes. They are also tissue-specific in that they rely on tissue-specific DNA-binding proteins for their activities. Sometimes a DNA element can act as either an enhancer or a silencer depending on what is bound to it.

SUMMARY

Eukaryotic nuclei contain three RNA polymerases that can be separated by ion-exchange chromatography. RNA polymerase I is found in the nucleolus; the other two polymerases are located in the nucleoplasm. The three nuclear RNA polymerases have different roles in transcription. Polymerase I makes a large precursor to the major rRNAs (5.8S, 18S, and 28S rRNAs in vertebrates). Polymerase II synthesizes hnRNAs, which are precursors to mRNAs. It also makes miRNA precursors and most small nuclear RNAs (snRNAs). Polymerase III makes the precursors to 5S rRNA, the tRNAs, and several other small cellular and viral RNAs.

The subunit structures of all three nuclear polymerases from several eukaryotes have been determined. All of these structures contain many subunits, including two large ones, with molecular masses greater than 100 kD. All eukaryotes seem to have at least some common subunits that are found in all three polymerases. The genes encoding all 12 RNA polymerase II subunits in yeast have been sequenced and subjected to mutation analysis. Three of the subunits resemble the core subunits of bacterial RNA polymerases in both structure and function, five are found in all three nuclear RNA polymerases, two are not required for activity, at least at normal temperatures, and two fall into none of these three categories.

Subunit IIa is the primary product of the RPB1 gene in yeast. It can be converted to IIb in vitro by proteolytic removal of the carboxyl-terminal domain (CTD), which is essentially a heptapeptide repeated over and over. Subunit IIa is converted in vivo to IIo by phosphorylating two serines within the CTD heptad. The enzyme (polymerase IIA) with the IIa subunit is the one that binds to the promoter; the enzyme (polymerase IIO) with the IIo subunit is the one involved in transcript elongation.

The structure of yeast pol II Δ4/7 reveals a deep cleft that can accept a DNA template. The catalytic center, containing a Mg^{2+} ion lies at the bottom of the cleft. A second Mg^{2+} ion is present in low concentration and presumably enters the enzyme bound to each substrate nucleotide.

The crystal structure of a transcription elongation complex involving yeast RNA polymerase II (lacking Rpb4/7) reveals that the clamp is indeed closed over the RNA–DNA hybrid in the enzyme's cleft, ensuring processivity of transcription. In addition, three loops of the clamp—the rudder, lid, and zipper—appear to play important roles in, respectively: initiating dissociation of the RNA–DNA hybrid, maintaining this dissociation, and maintaining dissociation of the template DNA. The active center of the enzyme lies at the end of pore 1, which appears to be the conduit for nucleotides to enter the enzyme and for extruded RNA to exit the enzyme during backtracking. A bridge helix lies adjacent to the active center, and flexing of this helix could play a role in translocation during transcription. The toxin α-amanitin appears to interfere with this flexing and thereby blocks translocation.

In moving through the entry pore toward the active site of RNA polymerase II, an incoming nucleotide first encounters the E (entry) site, where it is inverted relative to its position in the A site, the active site where phosphodiester bonds are formed. Two metal ions (Mg^{2+} or Mn^{2+}) are present at the active site. One is permanently bound to the enzyme and one enters the active site complexed to the incoming nucleotide. The trigger loop of Rpb1 positions the substrate for incorporation and discriminates against improper nucleotides.

The structure of the 12-subunit RNA polymerase II reveals that, with Rpb4/7 in place, the clamp is forced shut. Because initiation occurs with the 12-subunit enzyme, with its clamp shut, it appears that the promoter DNA must melt before the template DNA strand can descend into the enzyme's active site. It also appears that Rpb4/7 extends the dock region of the polymerase, making it easier for certain general transcription factors to bind, thereby facilitating transcription initiation.

Class II promoters may consist of a core promoter immediately surrounding the transcription start site, and a proximal promoter farther upstream. The core promoter may contain up to six conserved elements: the TFIIB recognition element (BRE), the TATA box, the initiator (Inr), the downstream core element (DCE), the motif ten element (MTE), and the downstream promoter element (DPE). At least one of these elements is missing in most promoters. Promoters for highly expressed specialized genes tend to have TATA boxes, but promoters for housekeeping genes tend to lack them.

Proximal promoter elements are usually found upstream of class II core promoters. They differ from the core promoter in that they bind to relatively gene-specific transcription factors. For example, GC boxes bind the transcription factor Sp1, while CCAAT boxes bind CTF. The proximal promoter elements, unlike the core promoter, can be orientation-independent, but they are relatively position-dependent, unlike classical enhancers.

Class I promoters are not well conserved in sequence from one species to another, but the general architecture of the promoter is well conserved. It consists of two elements: a core element surrounding the transcription start site, and an upstream promoter element (UPE) about 100 bp farther upstream. The spacing between these two elements is important.

RNA polymerase III transcribes a set of short genes. The classical class III genes (types I and II) have promoters that lie wholly within the genes. The internal promoter of the type I class III gene (the 5S rRNA gene) is split into three regions: box A, a short intermediate element, and box C. The internal promoters of the type II genes (e.g., the tRNA gene) are split into two parts: box A and box B.

Other class III genes called type III (e.g., 7SK, and U6 RNA genes) lack internal promoters altogether and contain promoters that strongly resemble class II promoters in that they lie in the 5′-flanking region and contain TATA boxes. The U1 and U6 snRNA genes have nonclassical class II and III promoters, respectively. The U1 snRNA promoter has an essential proximal sequence element (PSE) and a distal sequence element (DSE). The U6 snRNA promoter has a PSE, a DSE, and a TATA box.

Enhancers and silencers are position- and orientation-independent DNA elements that stimulate or depress, respectively, the transcription of associated genes. They are also tissue-specific in that they rely on tissue-specific DNA-binding proteins for their activities.

REVIEW QUESTIONS

1. Diagram the elution pattern of the eukaryotic nuclear RNA polymerases from DEAE-Sephadex chromatography. Show what you would expect if you assayed the same fractions in the presence of 1 μg/mL of α-amanitin.
2. Describe and give the results of an experiment that shows that polymerase I is located primarily in the nucleolus of the cell.
3. Describe and give the results of an experiment that shows that polymerase III makes tRNA and 5S rRNA.
4. How many subunits does yeast RNA polymerase II have? Which of these are "core" subunits? How many subunits are common to all three nuclear RNA polymerases?
5. Describe how epitope tagging can be used to purify polymerase II from yeast in one step.
6. Some preparations of polymerase II show three different forms of the largest subunit (RPB1). Give the names of these subunits and show their relative positions after SDS-PAGE. What are the differences among these subunits? Present evidence for these conclusions.
7. What is the structure of the CTD of RPB1?
8. Draw a rough diagram of the structure of yeast RNA polymerase II. Show where the DNA lies, and provide another piece of evidence that supports this location for DNA. Also, show the location of the active site.
9. How many Mg^{2+} ions are proposed to participate in catalysis at the active center of RNA polymerases? Why is one of these metal ions difficult to see in the crystal structure of yeast RNA polymerase II?
10. Cite evidence to support pore 1 as the likely exit point for RNA extrusion during polymerase II backtracking.
11. What is meant by the term "processive transcription?" What part of the polymerase II structure ensures processivity?
12. What is the probable function of the rudder of polymerase II?
13. What is the probable function of the bridge helix? What is the relationship of α-amanitin to this function?
14. What are the E site and A site of RNA polymerase II? What roles are they thought to play in nucleotide selection?
15. What role does the polymerase II trigger loop play in nucleotide selection? Illustrate with a schematic diagram of contacts to the base, sugar, and triphosphate.
16. What role does the Rpb4/7 complex play in opening or closing the clamp of RNA polymerase II? What evidence supports this role?
17. The 12-subunit RNA polymerase II interacts with promoter DNA. What implications does this have for the state of the promoter DNA with which the polymerase must interact?
18. Draw a diagram of a composite polymerase II promoter, showing all of the types of elements it could have.
19. What kinds of genes tend to have TATA boxes? What kinds of genes tend not to have them?
20. What is the probable relationship between TATA boxes and DPEs?
21. What are the two most likely effects of removing the TATA box from a class II promoter?
22. Describe the process of linker scanning. What kind of information does it give?
23. List two common proximal promoter elements of class II promoters. How do they differ from core promoter elements?
24. Diagram a typical class I promoter.
25. How were the elements of class I promoters discovered? Present experimental results.
26. Describe and give the results of an experiment that shows the importance of spacing between the elements of a class I promoter.
27. Compare and contrast (with diagrams) the classical and nonclassical class III promoters. Give an example of each.
28. Diagram the structures of the U1 and U6 snRNA promoters. Which RNA polymerase transcribes each? What is the effect of moving the TATA box from one of these promoters to the other? Why does this seem paradoxical?
29. Describe and give the results of an experiment that locates the 5′-border of the 5S rRNA gene's promoter.
30. Explain the fact that enhancer activity is tissue-specific.

ANALYTICAL QUESTIONS

1. Transcription of a class II gene starts at a guanosine 25 bp downstream of the last base of the TATA box. You delete 20 bp of DNA between this guanosine and the TATA box and transfect cells with this mutated DNA. Will transcription still start at the same guanosine? If not, where? How would you locate the transcription start site?
2. You suspect that a repeated sequence just upstream of a gene is acting as an enhancer. Describe and predict the results of an experiment you would run to test your hypothesis. Be sure your experiment shows that the sequence acts as an enhancer and not as a promoter element.
3. You are investigating a new class II promoter, but you can find no familiar sequences. Design an experiment to locate the promoter sequences, and show sample results.
4. Describe a primer extension assay you could use to define the 3′-end of the 5S rRNA promoter.

SUGGESTED READINGS

General References and Reviews

Corden, J.L. 1990. Tales of RNA polymerase II. *Trends in Biochemical Sciences* 15:383–87.

Klug, A. 2001. A marvelous machine for making messages. *Science* 292:1844–46.

Landick, R. 2004. Active-site dynamics in RNA polymerases. *Cell* 116:351–53.

Lee, T.I. and R.A. Young. 2000. Eukaryotic transcription. *Annual Review of Genetics* 34:77–137.

Paule, M.R. and R.J. White. 2000. Transcription by RNA polymerases I and III. *Nucleic Acids Research* 28:1283–98.

Sentenac, A. 1985. Eukaryotic RNA polymerases. *CRC Critical Reviews in Biochemistry* 18:31–90.

Woychik, N.A. and R.A. Young. 1990. RNA polymerase II: Subunit structure and function. *Trends in Biochemical Sciences* 15:347–51.

Research Articles

Benoist, C. and P. Chambon. 1981. In vivo sequence requirements of the SV40 early promoter region. *Nature* 290:304–10.

Bogenhagen, D.F., S. Sakonju, and D.D. Brown. 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.

Bushnell, D.A., P. Cramer, and R.D. Kornberg. 2002. Structural basis of transcription: α-amanitin–RNA polymerase cocrystal at 2.8 Å resolution. *Proceedings of the National Academy of Sciences USA* 99:1218–22.

Cramer, P., D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* 292:1863–76.

Das, G., D. Henning, D. Wright, and R. Reddy. 1988. Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. *EMBO Journal* 7:503–12.

Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33:717–28.

Gnatt, A.L., P. Cramer, J. Fu, D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292:1876–82.

Haltiner, M.M., S.T. Smale, and R. Tjian. 1986. Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. *Molecular and Cellular Biology* 6:227–35.

Kolodziej, P.A., N. Woychik, S.-M. Liao, and R. Young. 1990. RNA polymerase II subunit composition, stoichiometry, and phosphorylation. *Molecular and Cellular Biology* 10:1915–20.

Kutach, A.K. and J.T. Kadonaga. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Molecular and Cellular Biology* 20:4754–64.

Learned, R.M., T.K. Learned, M.M. Haltiner, and R.T. Tjian. 1986. Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* 45:847–57.

McKnight, S.L. and R. Kingsbury. 1982. Transcription control signals of a eukaryotic protein-coding gene. *Science* 217:316–24.

Murphy, S., C. Di Liegro, and M. Melli. 1987. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51:81–87.

Pieler, T., J. Hamm, and R.G. Roeder. 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.

Roeder, R.G. and W.J. Rutter. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224:234–37.

Roeder, R.G. and W.J. Rutter. 1970. Specific nucleolar and nucleoplasmic RNA polymerases. *Proceedings of the National Academy of Sciences USA* 65:675–82.

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

Sayre, M.H., H. Tschochner, and R.D. Kornberg. 1992. Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 267:23376–82.

Sklar, V.E.F., L.B. Schwartz, and R.G. Roeder. 1975. Distinct molecular structures of nuclear class I, II, and III DNA-dependent RNA polymerases. *Proceedings of the National Academy of Sciences USA* 72:348–52.

Smale, S.T. and D. Baltimore. 1989. The "initiator" as a transcription control element. *Cell* 57:103–13.

Ullu, E. and A.M. Weiner. 1985. Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature* 318:371–74.

Wang, D., D.A. Bushnell, K.D. Westover, C.D. Kaplan, and R.D. Kornberg. 2006. Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell* 127:941–954.

Weinman, R. and R.G. Roeder. 1974. "Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S rRNA genes." *Proceedings of the National Academy of Sciences USA* 71:1790–94.

Westover, K.D., D.A. Bushnell, and R.D. Kornberg. 2004. Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II. *Science* 303:1014–16.

Westover, K.D., D.A. Bushnell, and R.D. Kornberg. 2004. Structural basis of transcription: Nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119:481–89.

Woychik, N.A., S.M. Liao, P.A. Kolodziej, and R.A. Young. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes and Development* 4:313–23.

Woychik, N.A., et al. 1993. Yeast RNA polymerase II subunit RPB11 is related to a subunit shared by RNA polymerase I and III. *Gene Expression* 3:77–82.