



## Chapter 11

# Transcription in eukaryotes

... the modern researcher in transcriptional control has much to think about.

James T. Kadonaga, *Cell* (2004) 116:247.

### Outline

#### 11.1 Introduction

#### 11.2 Overview of transcriptional regulation

#### 11.3 Protein-coding gene regulatory elements

Structure and function of promoter elements  
Structure and function of long-range regulatory elements  
Focus box 11.1 Position effect and long-range regulatory elements

Disease box 11.1 Hispanic thalassemia and DNase I hypersensitive sites

#### 11.4 General (basal) transcription machinery

Components of the general transcription machinery  
Structure of RNA polymerase II  
General transcription factors and preinitiation complex formation  
Mediator: a molecular bridge  
Focus box 11.2 Is there a nuclear matrix?  
Focus box 11.3 Chromosomal territories and transcription factories

#### 11.5 Transcription factors

Transcription factors mediate gene-specific transcriptional activation or repression  
Transcription factors are modular proteins  
DNA-binding domain motifs  
Transactivation domain

Dimerization domain

Focus box 11.4 Homeoboxes and homeodomains

Disease box 11.2 Greig cephalopolysyndactyly syndrome and Sonic hedgehog signaling

Disease box 11.3 Defective histone acetyltransferases in Rubinstein-Taybi syndrome

#### 11.6 Transcriptional coactivators and corepressors

Chromatin modification complexes  
Linker histone variants  
Chromatin remodeling complexes  
Focus box 11.5 Is there a histone code?

#### 11.7 Transcription complex assembly: the enhanceosome model versus the “hit and run” model

Order of recruitment of various proteins that regulate transcription  
Enhanceosome model  
Hit and run model  
Merging of models

#### 11.8 Mechanism of RNA polymerase II transcription

Promoter clearance  
Elongation: polymerization of RNA  
Proofreading and backtracking

Transcription elongation through the nucleosomal barrier

Disease box 11.4 Defects in Elongator and familial dysautonomia

## 11.9 Nuclear import and export of proteins

Karyopherins

Nuclear localization sequences (NLSs)

Nuclear export sequences (NESs)

Nuclear import pathway

Nuclear export pathway

Focus box 11.6 The nuclear pore complex

Focus box 11.7 Characterization of the first nuclear localization sequence

## 11.10 Regulated nuclear import and signal transduction pathways

Regulated nuclear import of NF- $\kappa$ B

Regulated nuclear import of the glucocorticoid receptor

Chapter summary

Analytical questions

Suggestions for further reading

## 11.1 Introduction

Many eukaryotes are estimated to have 20,000–25,000 genes (see Table 16.4). Some of these are expressed (transcribed) in all cells all of the time, while others are expressed as cells enter a particular pathway of differentiation or as conditions in and around the cells change. In the early 1980s, transcription researchers primarily explored DNA–protein interactions *in vitro*. Research focused on the purification of sequence-specific DNA-binding proteins by affinity chromatography, analysis of the transcriptional activity of promoters by reporter gene assays, *in vitro* transcription assays that allowed the fractionation of the general transcription machinery, and assays such as electrophoretic gel mobility shift assays (EMSA) and DNase I footprinting for analysis of *cis*-acting DNA elements with *trans*-acting factors (see Chapter 9 for methods). By the late 1980s, many sequence-specific DNA-binding proteins had been identified, purified, and their genes cloned. Upon further study, it became clear that in addition to DNA–protein interactions, protein–protein interactions were of critical importance for regulating gene transcription. This insight was followed closely by the realization in the early 1990s that chromatin structure, nuclear architecture, and cellular compartmentalization must also be taken into account. Sections within this chapter will cover protein-coding gene regulatory elements, transcription factors and their DNA-binding motifs, the general transcription machinery and the mechanism of RNA polymerase II transcription, transcriptional coactivators and corepressors, including chromatin modification and remodeling complexes, and signal-mediated nuclear import and export of proteins involved in regulating gene transcription.

## 11.2 Overview of transcriptional regulation

The most important and widely used strategy for regulating gene expression is altering the rate of transcription of a gene. However, the control of gene expression can be exerted at many other levels, including processing of the RNA transcript, transport of RNA to the cytoplasm, translation of mRNA, and mRNA and protein stability. These additional levels of control are discussed in Chapters 13 and 14. There are also instances where genes are selectively amplified during development and, as a consequence, there is an increase in the amount of RNA transcript synthesized. The ribosomal RNA genes of *Xenopus* are an example of this form of gene regulation (see Fig. 6.17).

In this chapter, the regulation of transcription of protein-coding genes by RNA polymerase II (RNA pol II) will be highlighted. RNA pol II is located in the nucleoplasm and is responsible for transcription of the vast majority of genes including those encoding mRNA, small nucleolar RNAs (snoRNAs), some small nuclear RNAs (snRNAs), and microRNAs. Gene transcription is a remarkably complex process. The synthesis of tens of thousands of different eukaryotic mRNAs is carried out by RNA pol II. During the process of transcription, RNA pol II associates transiently not only with the template DNA but with many different proteins, including general transcription factors. The initiation step alone involves the assembly of dozens of factors to form a preinitiation complex. Transcription is mediated by the collective action of sequence-specific DNA-binding transcription factors along with the core RNA pol II transcriptional

machinery, an assortment of coregulators that bridge the DNA-binding factors to the transcriptional machinery, a number of chromatin remodeling factors that mobilize nucleosomes, and a variety of enzymes that catalyze covalent modification of histones and other proteins. Not surprisingly, the transcription literature is replete with a sometimes bewildering array of acronyms such as TBP, CBP, HDAC, LSD1, and SWI/SNF, to name a few (Table 11.1).

There are two other important eukaryotic polymerases – RNA polymerase I and RNA polymerase III (see Table 10.1). RNA polymerase I resides in the nucleolus and is responsible for synthesis of the large ribosomal RNA precursor. RNA polymerase III is also located in the nucleoplasm and is responsible for synthesis of transfer RNA (tRNA), 5S ribosomal RNA (rRNA), and some snRNAs. Plants have a fourth nuclear polymerase, named RNA polymerase IV, which is an RNA silencing-specific polymerase that mediates synthesis of small interfering RNAs (siRNAs) involved in heterochromatin formation (see Section 12.6). A full treatment of transcriptional regulation by all of the polymerases is beyond the scope of this chapter.

### 11.3 Protein-coding gene regulatory elements

Expression of protein-coding genes is mediated in part by a network of thousands of sequence-specific DNA-binding proteins called transcription factors. Transcription factors interpret the information present in gene promoters and other regulatory elements, and transmit the appropriate response to the RNA pol II transcriptional machinery. Information content at the genetic level is expanded by the great variety of regulatory DNA sequences and the complexity and diversity of the multiprotein complexes that regulate gene expression. Many different genes and many different types of cells in an organism share the same transcription factors. What turns on a particular gene in a particular cell is the unique combination of regulatory elements and the transcription factors that bind them.

Protein-coding sequences make up only a small fraction of a typical multicellular eukaryotic genome. For example, they account for less than 2% of the human genome. The typical eukaryotic protein-coding gene consists of a number of distinct transcriptional regulatory elements that are located immediately 5' of the transcription start site (termed +1). The regulatory regions of unicellular eukaryotes such as yeast are usually only composed of short sequences located adjacent to the core promoter (Fig. 11.1A). In contrast, the regulatory regions in multicellular eukaryotes are scattered over an average distance of 10 kb of genomic DNA with the transcribed DNA sequence only accounting for just 2 or 3 kb (Fig. 11.1B). Genes range in size from very small, such as a histone gene that is only 500 nt long with no introns, to very large. The largest known human gene encodes the protein dystrophin, which is missing or nonfunctional in the disease muscular dystrophy. The transcribed sequence is 2.5 million nucleotides in length, including 79 introns. It takes over 16 hours to produce a single transcript, of which more than 99% is removed during splicing to generate a mature mRNA.

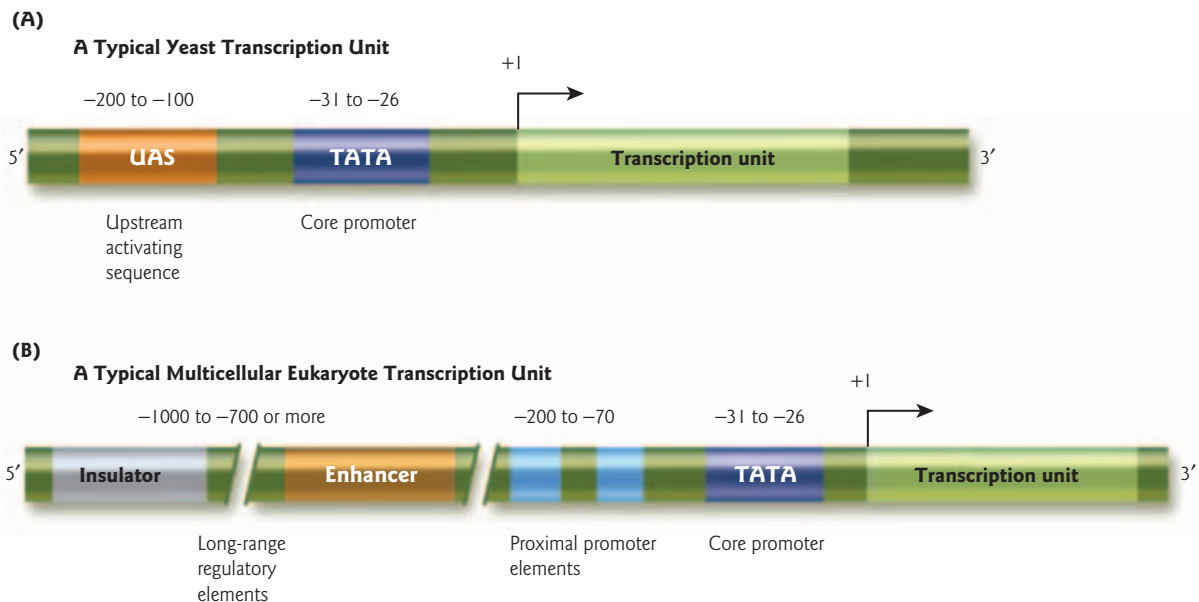
Gene regulatory elements are specific *cis*-acting DNA sequences that are recognized by *trans*-acting transcription factors (see Section 10.3 for more discussion of the terms “*cis*” and “*trans*”). *Cis*-regulatory elements in multicellular eukaryotes can be classified into two broad categories based on how close they are to the start of transcription: promoter elements and long-range regulatory elements. In comparing the regulatory region of a particular gene with another in multicellular eukaryotes, there will be variation in whether a particular element is present or absent, the number of distinct elements, their orientation relative to the transcriptional start site, and the distance between.

#### Structure and function of promoter elements

The “gene promoter” is loosely defined as the collection of *cis*-regulatory elements that are required for initiation of transcription or that increase the frequency of initiation only when positioned near the transcriptional start site. The gene promoter region includes the core promoter and proximal promoter elements. Proximal promoter elements are also sometimes designated as “upstream promoter elements” or “upstream regulatory elements.”

**Table 11.1 Proteins that regulate transcription.**

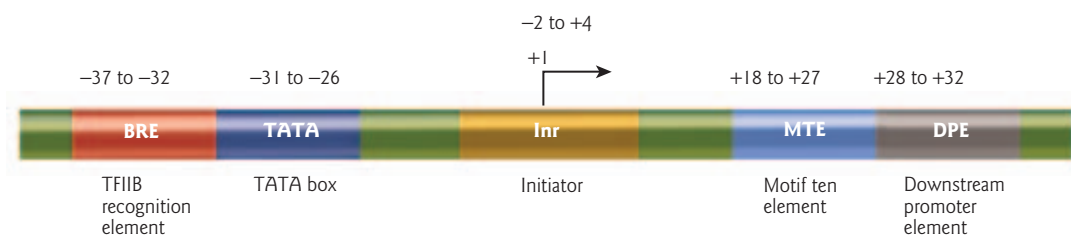
Category	Acronym	Derivation of name	Function
<b>Transcription factors (activators or repressors)</b>	Some examples mentioned in text:		
	CBF	<u>C</u> AAT binding factor	Binds CAAT box
	C/EBP	<u>C</u> AAT/ <u>e</u> nhancer-binding protein	Binds CAAT box
	CREB	<u>c</u> AMP response element-binding protein	Binds the cAMP response element
	CTCF	<u>C</u> <u>C</u> <u>C</u> <u>T</u> <u>C</u> -binding factor	Binds insulator element (CCCTC) and mediates enhancer blocking activity
	FOG-1	<u>F</u> riend of <u>G</u> ATA-1	Required for developmental expression of $\beta$ -globin genes
	GATA-1	GATA-binding protein	Required for developmental expression of $\beta$ -globin genes
	NF-E2	<u>N</u> uclear factor <u>e</u> rythoid-derived 2	Required for developmental expression of $\beta$ -globin genes
	NF- $\kappa$ B	<u>N</u> uclear factor of <u>k</u> appa light polypeptide enhancer in <u>B</u> cells	Central mediator of human stress and immune responses
	USF1, USF2	<u>U</u> pstream stimulatory factor 1 and 2	Bind insulator element, recruit chromatin-modifying enzymes
<b>General transcription machinery</b>	SATB1	<u>S</u> pecial <u>A</u> T-rich binding protein 1	Matrix attachment region (MAR) binding protein required for T-cell-specific gene regulation
	Sp1	<u>S</u> V40 early and late promoter-binding protein 1	Binds GC box
	RNA pol II (pol II, RNAPII)	RNA polymerase II	Catalysis of RNA synthesis
	General transcription factors:		
	TFIIB	<u>T</u> ranscription factor for RNA polymerase II <u>B</u>	Stabilization of TBP–DNA interactions, recruitment of RNA pol II–TFIIF, start site selection by RNA pol II
	TFIID:		
	TBP	<u>T</u> AATA-binding protein	Core promoter recognition, TFIIB recruitment
	TAF	<u>T</u> BPA-associated factor	Core promoter recognition/selectivity
	TFIIE	<u>T</u> ranscription factor for RNA polymerase II <u>E</u>	TFIIH recruitment
	TFIIF	<u>T</u> ranscription factor for RNA polymerase II <u>F</u>	Recruitment of RNA pol II to promoter DNA–TBP–TFIIB complex
<b>Coactivators and corepressors</b>	TFIIH	<u>T</u> ranscription factor for RNA polymerase II <u>H</u>	Promoter melting, helicase, RNA pol II CTD kinase
	Mediator	Mediator	Transduces regulatory information from activator and repressor proteins to RNA pol II
	Chromatin modification complexes:		
	HAT	<u>H</u> istone <u>a</u> cetyltransferase	Acetylates histones
	HDAC	<u>H</u> istone <u>d</u> eacetylase	Deacetylates histones
	CBP	<u>C</u> REB-binding protein	HAT activity
	HMT	<u>H</u> istone <u>m</u> ethyltransferase	Methylates histones
	LSD1	<u>L</u> ysine-specific <u>d</u> emethylase 1	Demethylates histones
	Chromatin remodeling complexes:		
	SWI/SNF	Mating-type <u>s</u> witching defective/ <u>s</u> ucrose nonfermenters	ATP-dependent chromatin remodeling (sliding and disassembly)
<b>Elongation factors</b>	ISWI	<u>I</u> mitation <u>S</u> wi2	ATP-dependent chromatin remodeling (sliding)
	SWR1	<u>S</u> wi2/ <u>S</u> nf2-related 1	ATP-dependent chromatin remodeling (histone replacement)
	FACT	<u>F</u> acilitates chromatin transcription	Transcription-dependent nucleosome alterations
	Elongator	Elongator	Exact function in elongation unknown
<b>Transcription factors (activators or repressors)</b>	TFIIS	<u>T</u> ranscription factor for RNA polymerase II <u>S</u>	Facilitates RNA pol II passage through regions that cause transcriptional arrest



**Figure 11.1 Comparison of a simple and complex RNA pol II transcription unit.** (A) A typical yeast (unicellular eukaryote) transcription unit. The start of transcription (+1) of the protein-coding gene (transcription unit) is indicated by an arrow. (B) A typical multicellular eukaryote transcription unit with clusters of proximal promoter elements and long-range regulatory elements located upstream from the core promoter (TATA). There is variation in whether a particular element is present or absent, the number of distinct elements, their orientation relative to the transcriptional start site, and the distance between them. Although the figure is drawn as a straight line, the binding of transcription factors to each other draws the regulatory DNA sequences into a loop.

### Core promoter elements

The core promoter is an approximately 60 bp DNA sequence overlapping the transcription start site (+1) that serves as the recognition site for RNA pol II and general transcription factors (see Section 11.4). Promoter elements become nonfunctional when moved even a short distance from the start of transcription or if their orientation is altered. The general transcription factor TFIID is responsible for the recognition of all known core promoter elements, with the exception of the BRE which is recognized by TFIIB. Some of the known core promoter elements are the TATA box, the initiator element (Inr), the TFIIB recognition element (BRE), the downstream promoter element (DPE), and the motif ten element (MTE) (Fig. 11.2,



**Figure 11.2 RNA pol II core promoter motifs.** Sequence elements that can contribute to basal transcription from the core promoter. A particular core promoter may contain some, all, or none of these motifs. The locations of the TFIIB recognition element (BRE), TATA box (TATA), initiator (Inr), motif ten element (MTE), and downstream promoter element (DPE) motifs are indicated relative to the start of transcription (+1).

**Table 11.2 Eukaryotic promoter elements.**

Promoter	Position	Transcription factor	Consensus sequence
<b>Upstream core promoter elements</b>			
TFIIB recognition element (BRE)	−37 to −32	TFIIB	(G/C)(G/C)(G/A)CGCC
TATA box	−31 to −26	TBP	TATA(A/T)AA(G/A)
Initiator (Inr)	−2 to +4	TAF1 (TAF <sub>II</sub> 250) TAF2 (TAF <sub>II</sub> 150)	PyPyA <sub>+1</sub> N(T/A)PyPy
<b>Downstream core promoter elements</b>			
Motif ten element (MTE)	+18 to +27	TFIID	C(G/A)A(A/G)C(G/C) (C/A/G)AACG(G/C)
Downstream promoter element (DPE)	+28 to +32	TAF9 (TAF <sub>II</sub> 40) TAF6 (TAF <sub>II</sub> 60)	(A/G)G(A/T)(C/T)(G/A/C)
<b>Proximal promoter elements</b>			
CAAT box	−200 to −70	CBF, NF1, C/EBP	CCAAT
GC box	−200 to −70	Sp1	GGGCGG

Most, but not all, CAAT and GC boxes are located between −200 and −70.

CBF, CAAT-binding protein; C/EBP, CAAT/enhancer-binding protein; N, any (A, T, C, or G); Py, pyrimidine (C or T).

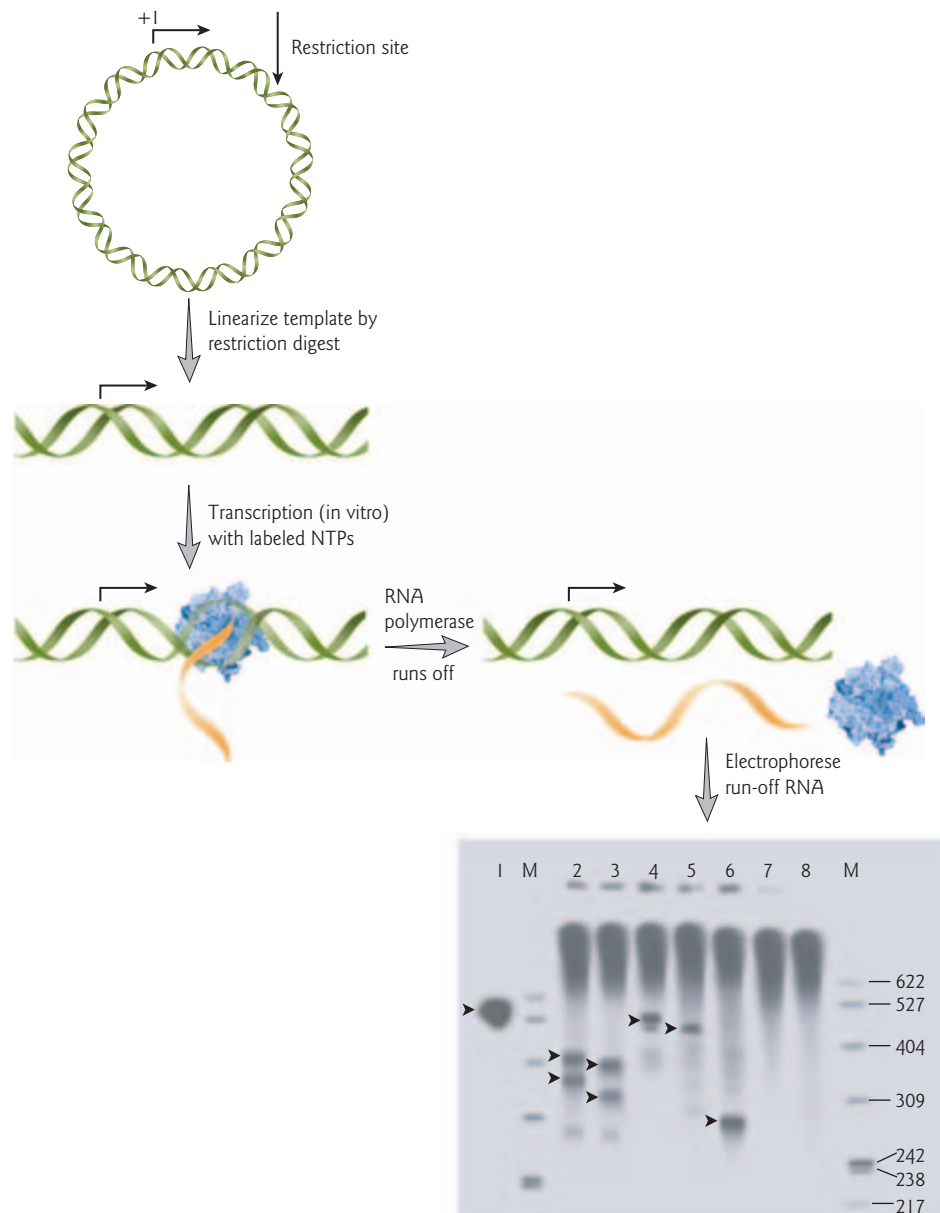
Table 11.2). Each of these sequence motifs is found in only a subset of core promoters. A particular core promoter may contain some, all, or none of these elements.

The TATA box (named for its consensus sequence of bases, TATAAA) was the first core promoter element identified in eukaryotic protein-coding genes. A key experiment by Pierre Chambon and colleagues demonstrated that a viral TATA box is both necessary and sufficient for specific initiation of transcription by RNA pol II. When they cloned a viral promoter into the plasmid pBR322, it was able to promote specific initiation of transcription *in vitro* (Fig. 11.3).

From early studies, it seemed that the TATA box was present in the majority of protein-coding genes. However, recent sequence database analysis of human genes found that TATA boxes are present in only 32% of potential core promoters. Thus, it has become increasingly important for molecular biologists to be aware of the full repertoire of possible promoter elements, and to continue the search for novel regulatory elements. The TATA box is the binding site for the TATA-binding protein (TBP), which is a major subunit of the TFIID complex (see Section 11.4). The TATA box can function in the absence of BRE, Inr, and DPE motifs. The Inr element was defined as a discrete core promoter element that is functionally similar to the TATA box. The Inr element is recognized by two other subunits of TFIID, TBP-associated factor 1 (TAF1) and TAF2 (TAF<sub>II</sub>250 and TAF<sub>II</sub>150 in the old TAF nomenclature). Inr can function independently of the TATA box, but in TATA-containing promoters, it acts synergistically to increase the efficiency of transcription initiation. Synergistic means that they act together, often to produce an effect greater than the sum of the two promoter elements acting separately. The Inr consensus sequence is shown in Table 11.2.

The DPE is a distinct seven nucleotide element that is conserved from *Drosophila* to humans. It functions in TATA-less promoters and is located about +30 relative to the transcription start site (see Fig. 11.2). The DPE consensus sequence is shown in Table 11.2. In contrast to the TATA box, the DPE motif requires the presence of an Inr. The DPE is bound by two specific subunits of the TFIID complex, TAF9 and TAF6 (TAF<sub>II</sub>40 and TAF<sub>II</sub>60, respectively, in the old TAF nomenclature). The recently identified MTE is located at positions +18 to +27 relative to the start of transcription (see Fig. 11.2 and Table 11.2). It promotes transcriptional activity and binding of TFIID in conjunction with the Inr. Although it can function independently of the TATA box or DPE, it exhibits strong synergism with both of these elements. Other downstream promoter motifs that contribute to transcriptional activity have been described that appear to be distinct from DPE and MTE. For example, the downstream core element (DCE) was first identified in the





**Figure 11.3 A TATA box-containing region promotes specific initiation of transcription *in vitro*.** A 21 or 43 bp DNA fragment containing the adenovirus type 2 major late promoter TATA box was cloned into plasmid pBR322. Plasmid “C” contains two copies of the 21 bp region, cloned in the same orientation. Plasmid “F” contains the 21 and 43 bp region cloned in the opposite orientation. Transcription was measured using an *in vitro* run-off assay, in which the plasmid template is linearized with a restriction endonuclease, then incubated in a cell-free extract with radiolabeled nucleoside triphosphates (NTPs). When RNA pol II (blue) reaches the end of the linear template, it falls off and releases the labeled run-off transcript (orange). Run-off transcripts are then separated by polyacrylamide gel electrophoresis and visualized by autoradiography. The size of the transcript corresponds to the distance between the start of transcription and the end of the template. The more actively the template is transcribed, the stronger the transcript signal. Lane 1 and lane M, size markers; lane 2, plasmid C cut with *Eco*RI; lane 3, plasmid C cut with *Hind*III; lane 4, plasmid F cut with *Eco*RI; lane 5, plasmid F cut with *Hind*III; lane 6, plasmid F cut with *Sal*I; lane 7, pBR322 wild-type plasmid (lacks a TATA box) cut with *Eco*RI; lane 8, pBR322 wild-type plasmid cut with *Sal*I. Arrowheads point to the run-off transcripts. No bands were produced by pBR322 lacking a TATA box. In all cases the TATA region directed RNA pol II to initiate about 30 bp downstream from the first T. Transcript sizes were similar to predicted sizes, indicating specific initiation from the TATA box (about 380 and 335 nt for plasmid C, and 515, 485, and 300 nt for plasmid F). (Reprinted from Sassone-Corsi, P., Corden, J., Kédinger, C., Chambon, P. 1981. Promotion of specific *in vitro* transcription by excised “TATA” box sequences inserted in a foreign nucleotide environment. *Nucleic Acids Research* 9:3941–3958, by permission of Oxford University Press).

human  $\beta$ -globin promoter. It consists of three sub-elements located at approximately +10, +20, and +30 of a subset of TATA-containing promoters. The DCE is bound by TAF1 and contributes to transcriptional activity of TATA-containing promoters.

### Proximal promoter elements

The regulation of TFIID binding to the core promoter element in yeast depends on an upstream activating sequence (UAS) located within a few hundred base pairs of the promoter (see Fig. 11.1A). The vast majority of yeast genes contain a single UAS, which is usually composed of two or three closely linked binding sites for one or two different transcription factors. In contrast, a typical multicellular eukaryote gene is likely to contain several proximal promoter elements. Promoter proximal elements are located just 5' of the core promoter and are usually within 70–200 bp upstream of the start of transcription. Recognition sites for transcription factors tend to be located in clusters. Examples include the CAAT box and the GC box (see Table 11.2). The CAAT box is a binding site for the CAAT-binding protein (CBF) and the CAAT/enhancer-binding protein (C/EBP). The GC box is a binding site for the transcription factor Sp1. Sp1 was initially identified as one of three components required for the transcription of SV40 early and late promoters. Promoter proximal elements increase the frequency of initiation of transcription, but only when positioned near the transcriptional start site. The transcription factors that bind promoter proximal elements do not always directly activate or repress transcription. Instead, they might serve as “tethering elements” that recruit long-range regulatory elements, such as enhancers, to the core promoter.

### Structure and function of long-range regulatory elements

Protein-coding genes of multicellular eukaryotes typically contain additional regulatory DNA sequences that can work over distances of 100 kb or more from the gene promoter. These long-range regulatory elements are instrumental in mediating the complex patterns of gene expression in different cells types during development. Such long-range regulation is not generally observed in yeast, although a few genes have regulatory sequences located further upstream than the UAS (e.g. silencers of the mating-type locus, see Section 12.7). The function of many long-range regulatory elements was confirmed by their effect on gene expression in transgenic animals. These elements tend to protect transgenes from the negative or positive influences exerted by chromatin at the site of integration (Focus box 11.1). Long-range regulatory elements in multicellular eukaryotes include enhancers and silencers, insulators, locus control regions (LCRs), and matrix attachment regions (MARs).

### Enhancers and silencers

A typical protein-coding gene is likely to contain several enhancers which act at a distance. These elements are usually 700–1000 bp or more away from the start of transcription. The hallmark of enhancers is that, unlike promoter elements, they can be downstream, upstream, or within an intron, and can function in either orientation relative to the promoter (Fig. 11.4). A typical enhancer is around 500 bp in length and contains in the order of 10 binding sites for several different transcription factors. Each enhancer is responsible for a subset of the total gene expression pattern. Enhancers increase gene promoter activity either in all tissues or in a regulated manner (i.e. they can be tissue-specific or developmental stage-specific). Similar elements that repress gene activity are called silencers.

### Insulators

Eukaryotic genomes are separated into gene-rich euchromatin and gene-poor, highly condensed heterochromatin. Because heterochromatin has a tendency to spread into neighboring DNA, natural barriers to spreading are critical when active genes are nearby. A mutation in *Drosophila* affecting a chromatin





Over the past 15–20 years of making transgenic animals and plants, a major problem has been a lack of expression of the transgene, or inappropriate expression (see Chapter 15). This has been attributed to “position effect.” Position effect is a phenomenon in which expression of the transgene is unpredictable; it varies with the chromosomal site of integration. Because integration is random when transgenic animals are made by pronuclear microinjection of foreign DNA (see Fig. 15.2), it is possible for the transgene to be integrated into either inactive or active chromatin. Because heterochromatin has a tendency to spread into neighboring DNA, transgenes that are integrated near heterochromatin tend to undergo inactivation. Traditionally, transgenes were constructed by fusing a complementary DNA (cDNA) coding for the protein of interest to a strong promoter. When a combination of long-range regulatory elements was included in the transgene construct, research found that position-independent expression units could be established, regardless of where the transgene integrated into the chromatin. Two classic examples of experiments showing how enhancers and MARs can protect transgenes from position effects are described below.

### **Intron enhancers contribute to tissue-specific gene expression**

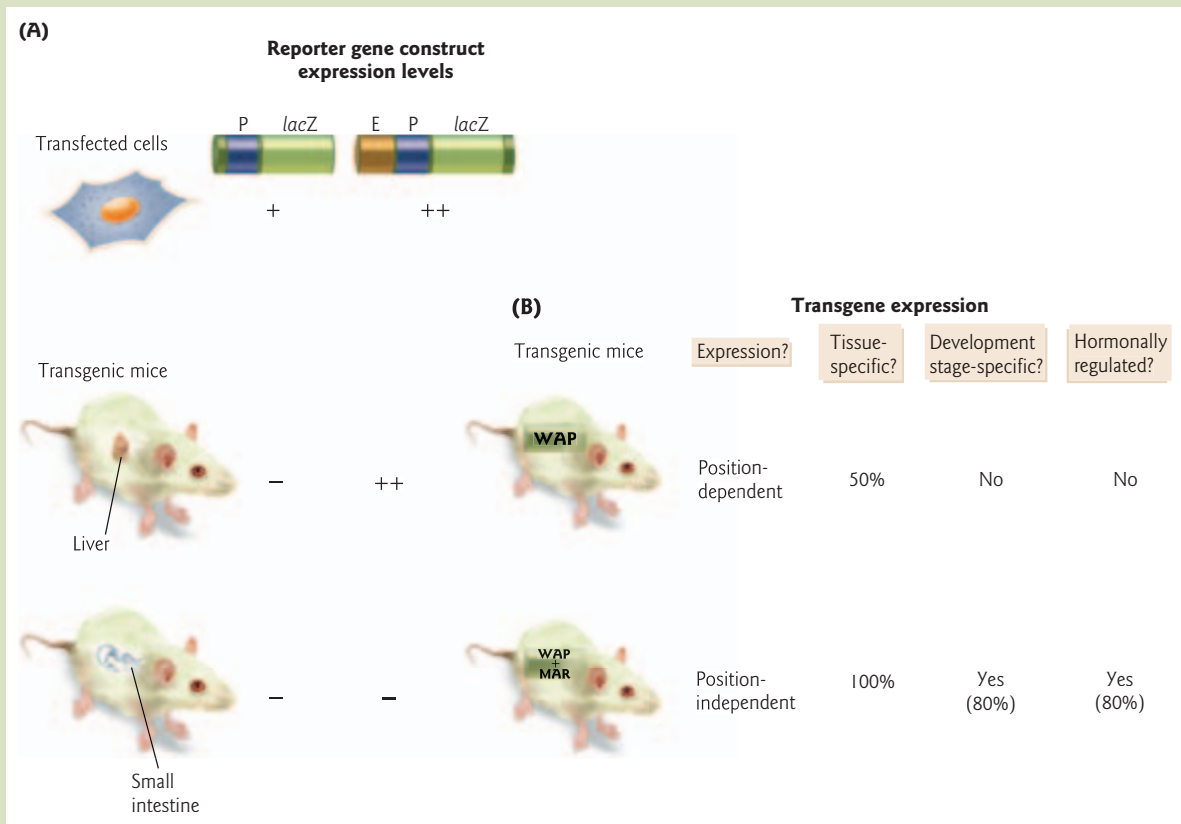
Introns were long considered to be “junk” DNA. We now know that they can include important coding DNA sequences (see Focus box 13.1), as well as regulatory elements such as enhancers. The importance of intron enhancers has been demonstrated for a number of genes *in vivo*. For example, Beatriz Levy-Wilson and colleagues showed that an enhancer located in the second intron of the human apolipoprotein B (*apoB*) gene is essential for tissue-specific gene expression. The gene product is a protein responsible for clearance of low-density lipoproteins (LDLs) by the LDL receptor and is involved in cholesterol homeostasis. The *apoB* gene is transcribed primarily in the liver and intestine in humans. In cell culture, a gene construct that contained only the *apoB* promoter linked to the *lacZ* reporter gene (see Section 9.3) was efficiently expressed. However, the addition of a sequence representing the second intron enhancer stimulated  $\beta$ -galactosidase activity 5–7-fold (Fig. 1A). In contrast, in transgenic mice there was no expression of the promoter-only construct. Integration of the construct was confirmed by Southern blot analysis (see Tool box 8.7) and transcription was assessed by RNase protection assay (see Fig. 9.8). In addition,  $\beta$ -galactosidase activity was assayed in

tissue sections of livers from transgenic and control mice. Further experiments showed that the second intron enhancer was absolutely required for expression of the reporter gene in the liver. Neither the promoter-only nor the promoter-enhancer construct were expressed in the small intestine, suggesting that additional tissue-specific regulatory elements are required.

### **MARs promote formation of independent loop domains**

The existence of matrix attachment regions (MARs) was considered to be without question; however, their biological significance was originally uncertain. Researchers thus set out to test whether MARs are essential for appropriately regulated gene expression *in vivo*. In one of the first studies of its kind, Lothar Hennighausen and colleagues examined transcriptional regulation of the whey acidic protein (WAP) gene, which codes for a major milk protein in mice. This gene provided an excellent model system because its expression is tissue-specific and developmentally and hormonally regulated. WAP is only expressed in the mammary gland during lactation, under control of the insulin–hydrocortisone–prolactin signaling pathway. Transgenic mice were generated with the 7 kb WAP-coding region and its associated promoter but no other flanking regions, or with the inclusion of chicken lysozyme gene MARs (Fig. 1B). The reason a chicken MAR was used was because at the time it was the best characterized MAR. A *HindIII* linker in the 5′ untranslated region was included to distinguish the transgene from the endogenous gene. Polymerase chain reaction (PCR) and Southern blot analysis were carried out to identify transgenic mice and to map the site of integration. Northern blot analysis was used to analyze gene transcription. The results showed that in the absence of the MARs, WAP mRNA expression was position-dependent. In other words, expression was unpredictable and depended on where the transgene was integrated. Expression was mammary-specific in 50% of the mouse lines, but it was hormone-independent and the levels were variable. In some mouse lines, WAP was activated early during pregnancy and then turned off during lactation. In contrast, in the presence of MARs, there was position-independent regulation. That is, when MARs were included, the transgene was unaffected by neighboring chromatin regardless of the site of integration. All transgenic mouse lines showed mammary-specific expression and four out of five lines showed accurate hormonal and developmental regulation.

This experiment and other similar ones (e.g. chicken lysozyme locus MARs have even been shown to reduce the

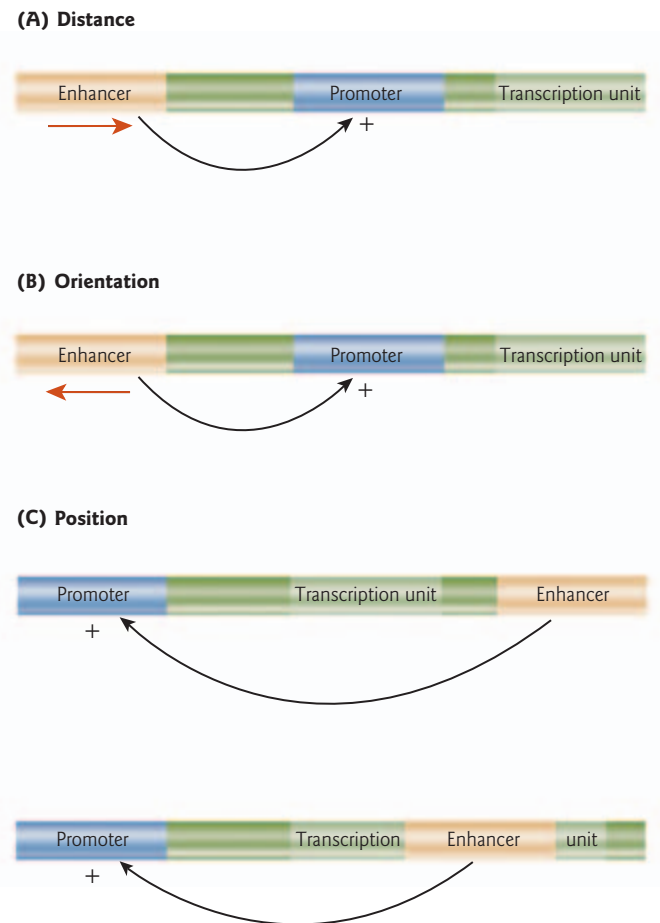


**Figure 1 Long-range regulatory elements can protect transgenes from position effect.** (A) Intron enhancer. In transiently transfected cells, a gene construct that contained only the apolipoprotein B gene promoter (P) linked to the *lacZ* reporter gene was efficiently expressed (indicated by “+” symbol) (see Fig. 9.1 and Section 9.3). The addition of a sequence representing the second intron enhancer (E) stimulated  $\beta$ -galactosidase activity 5–fold (++). In contrast, in transgenic mice (see Fig. 15.2) there was no expression of the promoter-only construct (indicated by “—” symbol). The second intron enhancer was absolutely required for expression of the reporter gene in the liver. Neither the promoter-only nor the promoter-enhancer construct were expressed in the small intestine. (Brand, M., Ranish, J.A., Kummer, N.T. et al. 1994. Sequences containing the second-intron enhancer are essential for transcription of the human apolipoprotein B gene in the livers of transgenic mice. *Molecular and Cellular Biology* 14:2243–2256. (B) Matrix attachment regions (MARs). Transgenic mice were generated with the whey acid protein (WAP) coding region and its associated promoter but no other flanking regions, or with the inclusion of chicken lysozyme gene MARs (WAP + MAR). In the absence of MARs, WAP mRNA expression was position-dependent. Expression was mammary-specific in 50% of the mouse lines, but expression was not developmentally or hormonally regulated. In the presence of MARs, there was position-independent regulation. All transgenic mouse lines showed mammary-specific expression and four out of five lines showed accurate hormonal and developmental regulation. (McKnight, R.A., Shamay, A., Sankaran, L., Wall, R.J., Hennighausen, L. 1992. Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proceedings of the National Academy of Sciences USA* 89:6943–6947.)

variability in transgene expression in rice plants) lend support for the following model of gene regulation. The association of MARs with the nuclear architecture allows formation of an independent DNA loop domain. This loop domain can adopt an altered chromatin structure distinct

from the structure of neighboring chromatin. In this altered configuration, the gene promoter and other regulatory elements become accessible to tissue-specific and/or developmental stage-specific transcription factors.

**Figure 11.4 Three key characteristics of an enhancer element.** An enhancer element can activate a promoter at a distance (A), in either orientation (B) or when positioned upstream, downstream, or within a transcription unit (C).

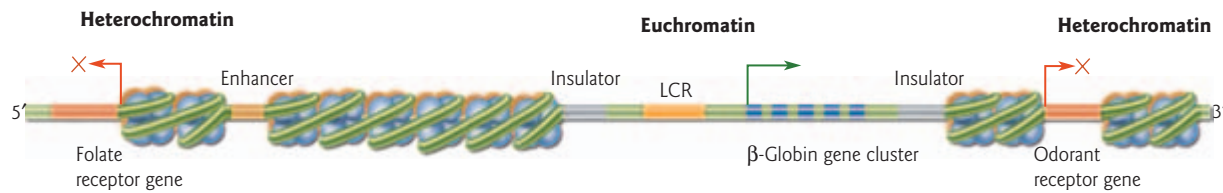


boundary was first observed by Ed Lewis in the 1970s. However, the general concept of boundary elements functioning as “insulators” was not fully established until the early 1990s. An insulator is a DNA sequence element, typically 300 bp to 2 kb in length, that has two distinct functions (Fig. 11.5):

- 1 Chromatin boundary marker:** an insulator marks the border between regions of heterochromatin and euchromatin.
- 2 Enhancer blocking activity:** an insulator prevents inappropriate cross-activation or repression of neighboring genes by blocking the action of enhancers and silencers.

Typically, insulators contain clustered binding sites for sequence-specific DNA-binding proteins. The exact molecular mechanism by which they block enhancers and silencers is not clear. One model proposes that insulators tether the DNA to subnuclear sites, forming loops that separate the promoter of one gene from the enhancer of another.

The vertebrate  $\beta$ -globin locus is an excellent model system for examining the interaction between insulator elements and chromatin structure. Gary Felsenfeld and co-workers first identified a chromatin boundary separating adjacent heterochromatin from  $\beta$ -globin genes within the locus. The boundary is located at a deoxyribonuclease I (DNase I) hypersensitive (HS) site called HS4. DNase I hypersensitive sites are a hallmark of active genes and regulatory elements (Disease box 11.1). Felsenfeld and colleagues went on to show that HS4 insulator also provides enhancer blocking activity from neighboring genes.



**Figure 11.5 Insulators function as chromatin boundary markers and have enhancer blocking activity.**

An insulator (gray) marks the border between regions of heterochromatin (depicted as nucleosome-bound DNA) and euchromatin. In erythrocytes it separates the actively transcribed  $\beta$ -globin genes (indicated by the green arrow) from the inactive folate receptor gene (indicated by the red X symbol). Another insulator prevents the inactive odorant receptor gene from cross-activation by the locus control region (LCR) (orange) upstream of the  $\beta$ -globin genes, and vice versa in other cell types. (Adapted from Bell, A.C., West, A.G., Felsenfeld, G. 2001. Insulators and boundaries. Versatile regulatory elements in the eukaryotic genome. *Science* 291:447–450.)

The insulator elements are recognized by at least three different DNA-binding proteins, CCCTC-binding factor (CTCF), and upstream stimulatory factor (USF) 1 and 2. CTCF mediates the enhancer blocking activity, while USF proteins bind to the insulator and recruit several chromatin-modifying enzymes (see Section 11.6).

### Locus control regions (LCRs)

LCRs are DNA sequences that organize and maintain a functional domain of active chromatin and enhance the transcription of downstream genes. Although sometimes referred to as “enhancers” of transcription, LCRs, unlike classic enhancer elements, operate in an orientation-dependent manner. The prototype LCR was characterized in the mid-1980s as a cluster of DNase I-hypersensitive sites upstream of the  $\beta$ -globin gene cluster (Fig. 11.6A). At the time, DNase I-hypersensitive sites were known to be important elements in the control of chromatin structure and transcriptional activity (Disease box 11.1). This series of hypersensitive sites became known collectively first as a “locus activation region” and later as the “locus control region.” Subsequently, LCRs have been shown to be present in other loci, including gene clusters encoding the  $\alpha$ -globins, visual pigments, major histocompatibility proteins, human growth hormones, serpins (a family of structurally related proteins that inhibit proteases), and T-helper type 2 cytokines (involved in the immune response).

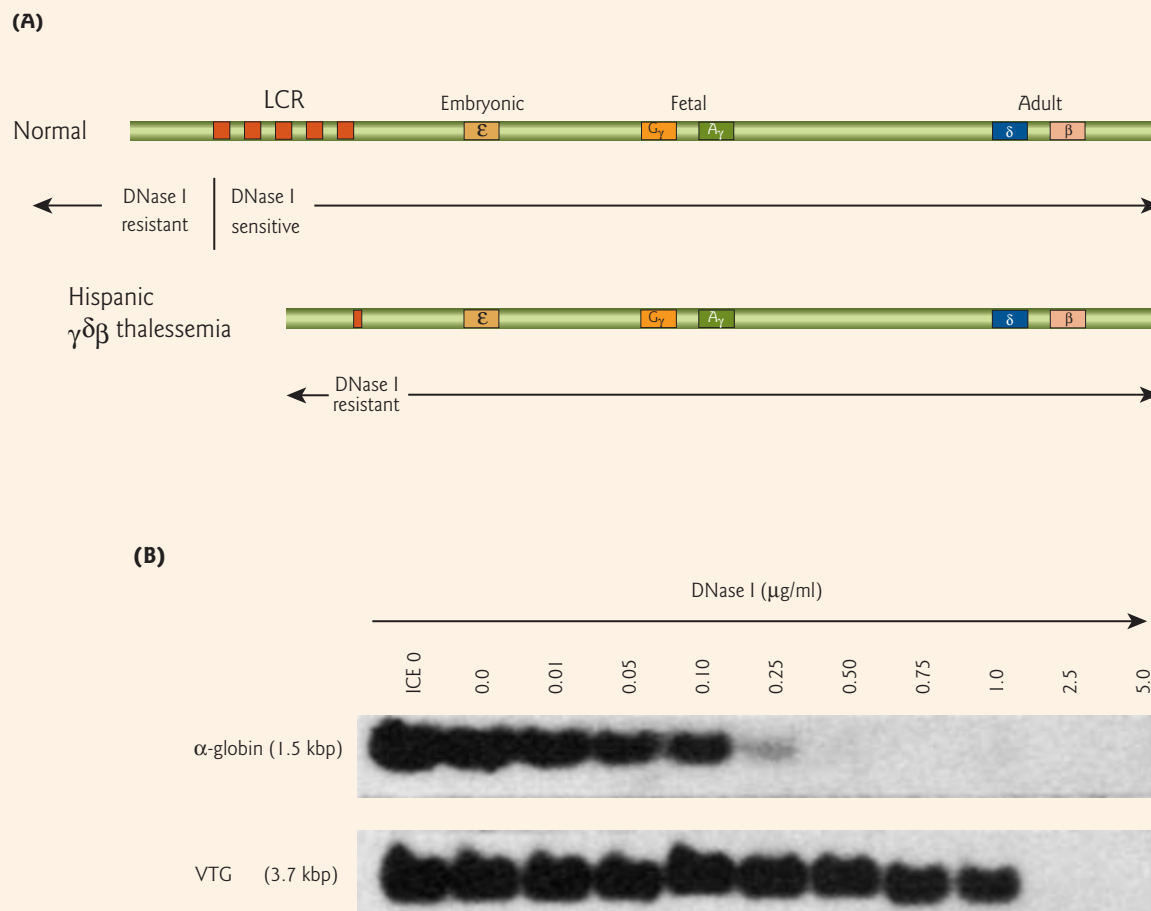
**Beta-globin gene LCR is required for high-level transcription** Hemoglobin is the iron-containing oxygen transport metalloprotein in the red blood cells of mammals and other animals. In adult humans, the most common hemoglobin is a tetramer composed of two  $\alpha$ -like globin polypeptides and two  $\beta$ -like polypeptides plus four heme groups (an organic molecule with an iron atom) (see Fig. 5.10). Given their critical function, it is not surprising that the  $\alpha$ - and  $\beta$ -globin genes are highly regulated. In particular, the LCR of the  $\beta$ -like globin gene cluster provides an excellent illustration of the complexity of regulatory regions (Fig. 11.6A). The  $\beta$ -like globin-coding regions are each 2–3 kb in size and the entire cluster spans approximately 100 kb. The genes are expressed in erythroid cells in a tissue- and developmental stage-specific manner: the epsilon ( $\epsilon$ ) globin gene is activated in the embryonic stage, the gamma ( $\gamma$ ) globin is activated in the fetal stage, and the  $\beta$ -globin gene is expressed in adults. Physiological levels of expression of each of these genes can be achieved only when they are downstream of the LCR. The DNase hypersensitive sites contain clusters of transcription factor-binding sites and interact via extensive protein–DNA and protein–protein interactions.

Early studies of  $\beta$ -globin gene expression *in vivo* were often inconclusive. Proper developmental regulation and high-level expression could not be achieved coordinately in transgenic mice carrying an artificial



## DISEASE BOX 11.1

## Hispanic thalassemia and DNase I hypersensitive sites



**Figure 1 Hispanic thalassemia and DNase I sensitivity.** (A) An approximately 35 kb deletion of the  $\beta$ -globin LCR causes Hispanic  $\gamma\delta\beta$ -thalassemia. The Hispanic locus is transcriptionally silent and the entire gene cluster is DNase I-resistant. In normal individuals the locus is transcriptionally active; chromatin upstream of the LCR is DNase I-resistant, whereas downstream of the LCR it is DNase I-sensitive. (B) Method for showing that transcriptionally active genes are more susceptible than inactive genes to DNase I digestion. Chick embryo erythroblasts at 15 days actively synthesize  $\alpha$ -globin, but not vitellogenin. Nuclei were isolated and exposed to increasing concentrations of DNase I. The nuclear DNA was extracted and digested with the restriction endonuclease *Bam*HI, which cleaves the  $\alpha$ -globin gene resulting in a 1.5 kb fragment and the vitellogenin gene resulting in a 3.7 kb fragment. The digested DNA was analyzed by Southern blot with a probe of labeled  $\alpha$ -globin DNA. The transcriptionally active  $\alpha$ -globin DNA from the 15-day erythroblasts was sensitive to DNase I digestion, indicated by the absence of the 1.5 kb band at higher nuclease concentrations. In contrast, the inactive vitellogenin DNA (VTG) was resistant to DNase I digestion, indicated by the presence of the 3.7 kb band at higher nuclease concentrations. (Reprinted with permission from Conklin, K.F., Groudine, M. 1986. Varied interactions between proviruses and adjacent host chromatin. *Molecular and Cellular Biology* 6:3999–4007. Copyright © 1986 American Society for Microbiology.)

## Hispanic thalassemia and DNase I hypersensitive sites

DISEASE BOX 11.1



$\gamma\delta\beta$ -Thalassemia is a rare disorder characterized by partial or complete deletions of the most 5' sequences of the  $\beta$ -like globin gene cluster, leading to reduced amounts of hemoglobin in the blood. Usually babies are diagnosed with the disease between the ages of 6 and 18 months. Depending on the extent of the deletion, and whether a patient is heterozygous or homozygous for the mutation, symptoms range from severe to more mild forms of anemia. Regular transfusion with red blood cells may be necessary to sustain life. Analysis of patients with this disease has led to significant advances in understanding of the locus control region (LCR) of the  $\beta$ -globin gene locus.

### Hispanic thalassemia

In 1989, a naturally occurring ~35 kb deletion of the LCR was found in a Hispanic patient with a form of  $\gamma\delta\beta$ -thalassemia now called "Hispanic thalassemia." The LCR deletion was shown to result in drastic changes in activity of the  $\beta$ -globin locus. It quickly became apparent that deletions of the LCR in the  $\beta$ -globin gene cluster result in silencing of the genes, even though the genes themselves are intact. Analysis of the wild-type and Hispanic deletion alleles in mouse erythroid cells (following transfer of these chromosomes from the heterozygous patient into the cells) showed that the Hispanic locus was transcriptionally silent and the entire region of the  $\beta$ -like globin gene cluster was DNase I-resistant (Fig. 1). These findings led researchers to propose that normally the LCR maintains an "open" chromatin structure and enhances transcription by

establishing an independent domain. Deletion of the LCR leads to "closed" chromatin and inactive  $\beta$ -like globin genes.

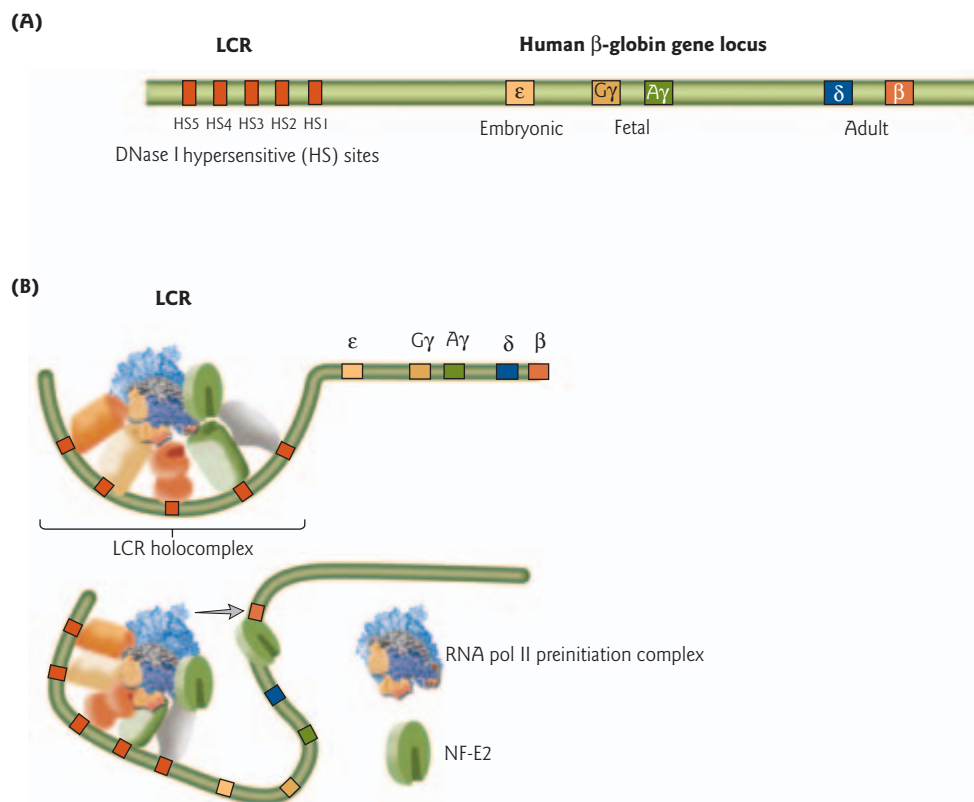
### Analysis of DNase I sensitivity

When chromatin is digested *in situ* with a low concentration of DNase I, certain regions are particularly sensitive to the nuclease. Such DNase I sensitivity is one feature of genes that are able to be transcribed. The nuclease introduces double-strand breaks in transcriptionally active chromatin over 100 times more frequently than in inactive chromatin. In addition to this general sensitivity to nucleases, there are also short DNA sequences (100–200 bp) called DNase I hypersensitive sites. These sites are the first place DNase I introduces a double-strand break in chromatin, and are > 2 orders of magnitude more accessible to cleavage compared with neighboring active chromatin. These sites are typically composed of clusters of recognition sites for sequence-specific DNA-binding proteins. DNase I hypersensitive sites are not necessarily nucleosome free. They may represent the stable association of a transcription factor or complex on the surface of the nucleosome. Figure 1 shows the relative levels of nuclease sensitivity in chromatin from 15-day-old chicken embryo erythrocytes. A comparison was made between the rates of DNase I digestion of the  $\alpha$ -globin gene which is expressed in erythrocytes, and the vitellogenin gene which is only expressed in the liver. Results show that the inactive vitellogenin gene is DNase I-resistant and the transcriptionally active globin genes are DNase I-sensitive.

construct, such as a plasmid-based  $\beta$ -globin gene construct. Advances in understanding LCR function came with the development of yeast artificial chromosome (YAC) vectors (see Section 8.4). Unlike plasmid vectors, which have an insert size limit of less than 10 kb, YACs can stably maintain a large enough insert of foreign DNA to encompass the entire  $\beta$ -globin gene locus in its natural configuration, which spans about 200 kb. Transgenic mouse experiments revealed that the LCR is required for high-level transcription of all the  $\beta$ -like globin genes, but the regulation of chromatin "loop" formation is the main mechanism controlling developmental expression of the  $\beta$ -globin genes.

Both developmentally proper regulation and physiological levels of expression of the  $\beta$ -globin genes can be recaptured in transgenic mice carrying a YAC construct. For example, a transgenic mouse line was constructed carrying a  $\beta$ -globin locus YAC that lacked the LCR. RNase protection assays showed no detectable levels of  $\epsilon$ -,  $\gamma$ -, or  $\beta$ -globin gene transcripts at any stage of development. These results demonstrated conclusively that the LCR is a minimum requirement for globin gene expression. Additional *in vivo* mutagenesis studies have shown the developmental regulation of the  $\beta$ -like globin genes is mediated





**Figure 11.6 The human  $\beta$ -like globin gene locus.** (A) Diagrammatic representation of the human  $\beta$ -like globin gene locus on chromosome 11, which encodes embryonic  $\epsilon$ -globin, the two fetal  $\gamma$ -globins, and the adult  $\delta$ - and  $\beta$ -globins. The locus control region (LCR) upstream of the  $\epsilon$ -globin gene has five DNase I hypersensitive (HS) sites (HS1–HS5) separated from each other by 2–4 kb. (B) Model for transcription complex recruitment. The general transcription machinery (RNA pol II and the preinitiation complex) and other transcriptional regulatory proteins are recruited to the LCR to form a “holocomplex.” A developmental stage-specific chromatin loop forms and transcription complexes are then transferred from the LCR to the appropriate globin gene promoter. The transfer is facilitated by the transcription factors NF-E2, GATA-1, and FOG-1. (Adapted with permission of The American Society for Biochemistry and Molecular Biology, Inc. from Vieira, K.F., Levings, P.P., Hill, M.A., Cruselle, V.J., Kang, S.H.L., Engel, J.D., Bungert, J. 2004. Recruitment of transcription complexes to the  $\beta$ -globin gene locus *in vivo* and *in vitro*. *Journal of Biological Chemistry* 279:50350–50357; permission conveyed through the Copyright Clearance Center, Inc.)

by the regulatory elements within their individual promoters. Targeted deletions in mice reveal an absolute requirement of the LCR for high-level transcription of all the  $\beta$ -like globin genes, but the LCR interacts with only one gene promoter at any one time.

**Chromatin “loop” formation controls developmental expression of the  $\beta$ -globin genes** Chromatin immunoprecipitation (ChIP) assays (see Fig. 9.15E) were used to demonstrate that RNA pol II is first recruited to LCR DNase I hypersensitive sites *in vivo*. The transfer of RNA pol II from the LCR to the  $\beta$ -globin gene promoter is stimulated by the erythroid transcription factor NF-E2 (Fig. 11.6B). Researchers showed that the transcription factor GATA-1 and its cofactor FOG-1 (friend of GATA-1) are required for the physical interaction between the  $\beta$ -globin LCR and the  $\beta$ -globin promoter. GATA-1 has a zinc finger DNA-binding motif (see Section 11.5) and binds to the DNA sequence 5′-GATA-3′. ChIP assays were used

to show that direct interaction with FOG-1 is required for GATA to induce formation of a tissue-specific chromatin “loop.” In this context, the term “loop” means a chromatin conformation where two distant regions of DNA located in *cis* along the chromatin are physically close to one another, but not to intervening DNA sequences. Because GATA-1 induced loop formation correlates with the onset of  $\beta$ -globin gene transcription, regulation of loop formation is thought to be the main mechanism controlling developmental expression of the  $\beta$ -globin genes. This model provides an explanation of how the LCR can enhance the rate of transcription over such large distances. As we saw in Section 10.6, such DNA looping is an equally important transcriptional regulatory mechanism in bacteria.

### Matrix attachment regions (MARs)

Much of the initial work on the basic molecular mechanisms of gene expression was done in simple test tube systems. However, it is now becoming clear that the three-dimensional organization of chromatin within the cell nucleus plays a central role in transcriptional control. There is increasing evidence that eukaryotic chromatin is organized as independent loops. Following histone extraction, these loops can be visualized as a DNA halo anchored to a densely stained matrix or chromosomal scaffold (Focus boxes 11.2 and 11.3). The formation of each loop is dependent on specific DNA sequence elements that are scattered throughout the genome at 5–200 kb intervals. These DNA sequences are termed either scaffold attachment regions (SARs) when prepared from metaphase cells, or matrix attachment regions (MARs) when prepared from interphase cells. MARs are thought to organize the genome into approximately 60,000 chromatin loops with an average loop size of 70 kb. Active genes tend to be part of looped domains as small as 4 kb, whereas inactive regions of chromatin are associated with larger domains of up to 200 kb.

Greater than 70% of characterized MARs are AT rich. The particular mode of MAR–matrix interaction indicates that binding is not directly correlated to the primary DNA sequence. Instead, the secondary structure-forming potential of DNA that tends to unwind in the AT-rich patches is of greater importance. In addition some MARs are GT rich and have the potential to form Z-DNA (see Fig. 2.8). MARs are typically located near enhancers in 5′ and 3′ flanking sequences. They are thought to confer tissue specificity and developmental control of gene expression by recruiting transcription factors and providing a “landing platform” for several chromatin-remodeling enzymes. Some MARs include recognition sites for topoisomerase II.

A model for the organization of transcription in active loop domains which localize transcription factors and actively transcribed genes is shown in Fig. 11.7. Two types of nuclear matrix-binding sites are proposed to exist within the loops, structural and functional MARs (also known as constitutive and facultative, respectively). Structural MARs serve as anchors, whereas functional MARs are more dynamic and help to bring genes onto the nuclear matrix. MARs reversibly associate with ubiquitous factors. These contacts can be altered by specific interactions with components of enhancers and LCRs. For example, SATB1 (special AT-rich-binding protein 1) is one of the best-characterized MAR-binding proteins. SATB1 is preferentially expressed in thymocytes, the precursors of T cells in the immune system. The protein binds to the base of the chromatin loop and is thought to play a key role in T cell-specific gene regulation.

## 11.4 General (basal) transcription machinery

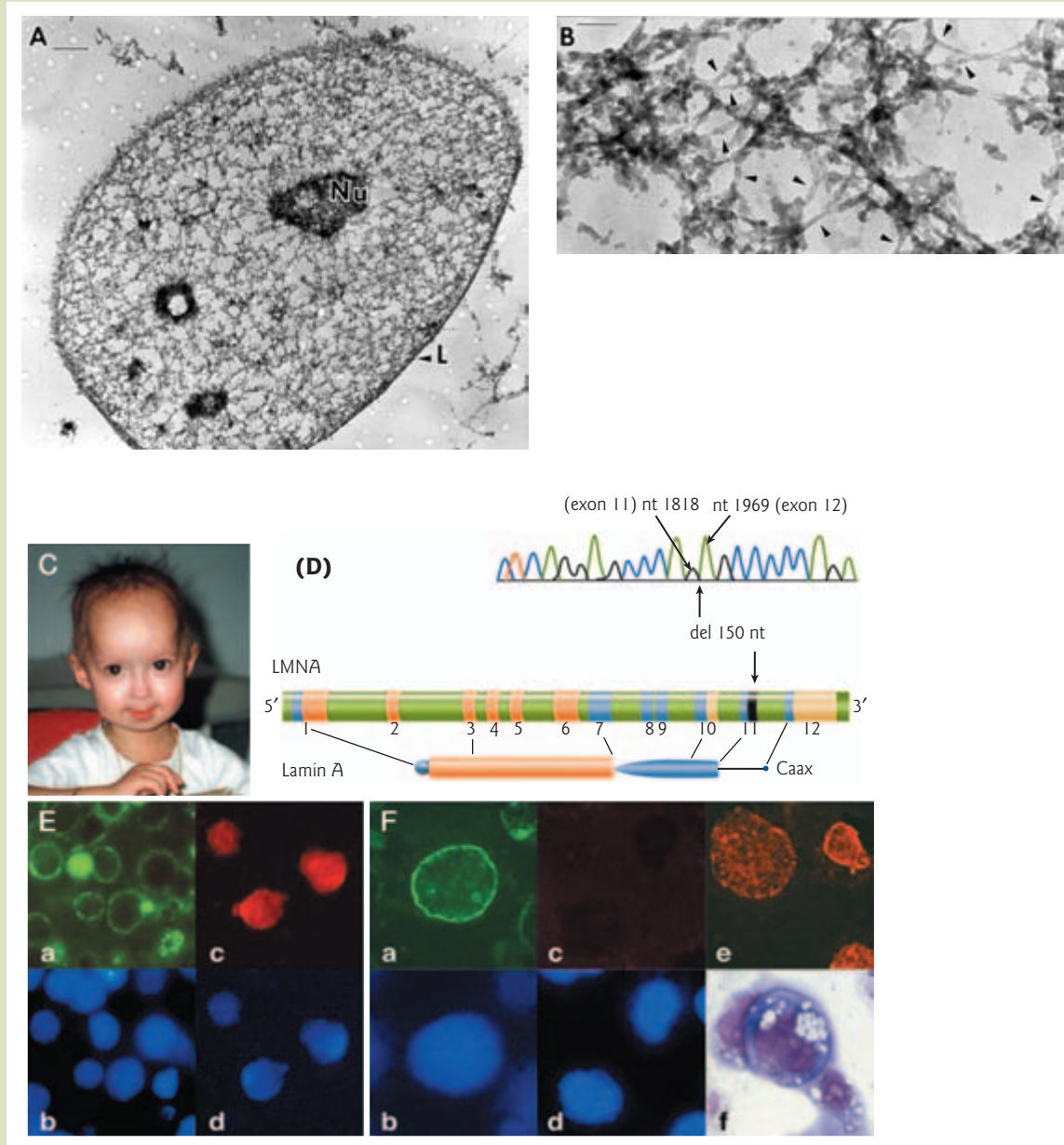
Five to 10% of the total coding capacity of the genome of multicellular eukaryotes is dedicated to proteins that regulate transcription. The yeast genome encodes a total of approximately 300 proteins involved in the regulation of transcription, while there may be as many as 1000 in *Drosophila* and 3000 in humans. These proteins fall into three major classes (see Table 11.1):

- 1 The general (basal) transcription machinery (this section): general, but diverse, components of large multiprotein RNA polymerase machines required for promoter recognition and the catalysis of RNA synthesis.



## FOCUS BOX 11.2

## Is there a nuclear matrix?



**Figure 1 The nuclear matrix.** (A) Resinless section electron micrograph of the core filaments of a CaSki cell (human cell line) nuclear matrix. Soluble proteins and chromatin have been removed from this nucleus. The core filament network, shown at high magnification, is connected to the nuclear lamina (L). Remnants of nucleoli (Nu) remain and are connected to fibers of the internal nuclear matrix. Bar, 1.0  $\mu\text{m}$ . (B) Higher magnification showing the fibrous network structure in more detail. The underlying 10-nm filaments are seen most clearly when they are free of covering material (arrowheads). Bar, 100 nm. (Reprinted with permission from:

## Is there a nuclear matrix?

FOCUS BOX 11.2



Prior to the mid-1970s, the nucleus was viewed by many scientists as “a bag of chromatin floating in a sea of nucleoplasm.” This was despite many observations to the contrary. For example, as early as 1925, Edmund B. Wilson in his classic textbook *The Cell in Development and Heredity* described the nucleus as containing a “nuclear framework” that was a “net-like or sponge-like reticulum,” and in 1948 researchers observed that extraction of nuclei with high salt solutions produced a residual structure. In 1974, Ronald Berezny and D.S. Coffey also extracted a salt-insoluble residual structure from nuclei that they called the “nuclear matrix.”

At first the general scientific community was skeptical that this meshwork was an artefact because of the harsh methods used to extract it from nuclei. After all, the nuclear matrix is operationally defined as “a branched meshwork of insoluble filamentous proteins within the nucleus that remains after digestion with high salt, nucleases, and detergent.” In the mid-1980s techniques for visualizing the meshwork in whole cells were developed that provided evidence that such an architectural component exists (Fig. 1). At the same time, observations highlighted its dynamic nature. Because of its dynamic elusive qualities, this structure has been referred to at various times as the nuclear skeleton, nuclear scaffold, or even “chickenwire.” The current preferred term is “matrix” (a dimensional field of variables) over the alternative terms that imply rigidity.

### What does the nuclear matrix do?

The nuclear matrix is proposed to serve as a structural organizer within the cell nucleus. For example, direct interaction of matrix attachment regions (MARs) with the nuclear matrix is proposed to organize chromatin into loop domains and to maintain chromosomal territories (see Focus box 11.3). Active genes are found associated with the nuclear matrix only in cell types in which they are expressed. In cell types where they are not expressed, these genes no longer associate with the nuclear matrix.

### What are the components of the nuclear matrix?

There are over 200 types of proteins associated with the nuclear matrix, of which most have not been characterized. What forms the framework for the branching filaments remains unknown. Some nuclear matrix proteins are common to all cell types and others are tissue-specific. For example, steroid receptors, such as the glucocorticoid receptor, have been shown to be associated with the nuclear matrix in hormone-responsive cells. General components of the matrix include the heterogeneous nuclear ribonucleoprotein (hnRNP) complex proteins and the nuclear lamins; hnRNP proteins are involved in transcription, transport, and processing of hnRNA.

The nuclear lamina is a protein meshwork underlying the nuclear membrane that is primarily composed of the intermediate filament proteins lamins A, B, and C. Internal

Nickerson, J.A., Krockmalnic, G., Wan, K.M., Penman, S. 1997. *Proceedings of the National Academy of Sciences USA* 94:4446–4450. Copyright ©1997 National Academy of Sciences, USA. Photographs courtesy of Jeffrey Nickerson, University of Massachusetts Medical Center.) (C–F) Lamin A truncation in Hutchinson–Gilford progeria, a premature aging syndrome. (C) Hutchinson–Gilford progeria affecting a 6-year-old female. (D) Schematic representation of the *LMNA* gene, which encodes both lamin A and lamin C, and of the lamin A protein, correlated by blue (globular domains) and red (rod domains) colors. The deleted *LMNA* transcript junction sequence is shown. The 150 bp deletion (indicated by a black bar) extends from G1819 to the end of exon 11. The CaaX (cysteine–aliphatic–aliphatic–any amino acid) sequence mediates farnesylation of the protein. (E) Nuclei from normal lymphocytes: (a) detection of lamin A/C and (c) lamin A by immunostaining; (b and d) DAPI staining of DNA. (F) Nuclei from the patient: (a) detection of lamin A/C by immunostaining; (c) absence of lamin A; (b and d) DAPI staining of DNA; (e) lamin B1 localizes both at the nuclear envelope (normal distribution) and nucleoplasm (abnormal distribution); (f) Giemsa staining shows nuclear deformities and cytoplasmic vacuoles. Width of fluorescent images, 80  $\mu\text{m}$ ; width of brightfield images, 120  $\mu\text{m}$ . (Parts C–F reprinted with permission from: De Sandre-Giovannoli, A., Bernard, R., Cau, P. et al. 2003. Lamin A truncation in Hutchinson–Gilford progeria. *Science* 300:2055. Copyright © 2003 AAAS).

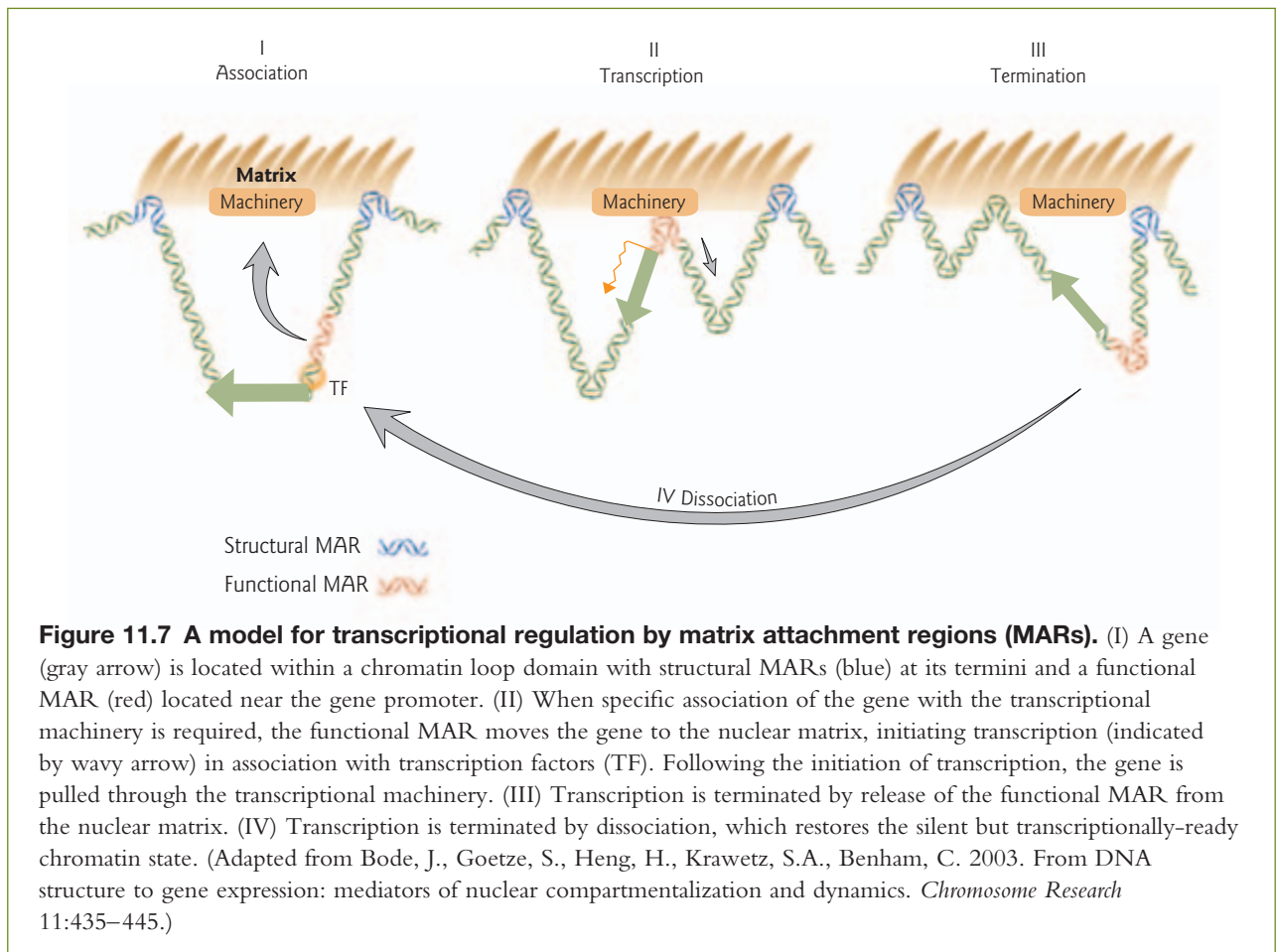
lamins form a “veil” that appears to branch throughout the interior of the nucleus (Fig. 1A,B). The importance of the lamina is highlighted by the wide variety of human disorders resulting from mutations in the *LMNA* locus which encodes lamin A/C proteins. For example, a splicing mutation in the lamin A gene leads to expression of a truncated form of the protein and causes Hutchinson–Gilford progeria syndrome – a premature aging syndrome in which patients have an average life expectancy of ~13 years. Patient cells have altered nuclear sizes and shapes, with disrupted nuclear membranes and extruded chromatin (Fig. 1C–F).

There are other filament-forming protein families present in the cell nucleus, but none seems to account for the long-range filament system visualized after high salt extraction. Actin was considered a candidate for the meshwork; however, although  $\beta$ -actin is present in the nucleus, it remains in monomer form. Two nuclear pore

complex-associated proteins, Nup153 and Tpr, are organized into filaments that may play a role in mRNA export. But these filaments only extend 100–350 nm into the nucleus, ruling them out as the component of the long-range filament system.

### Is there a nuclear matrix?

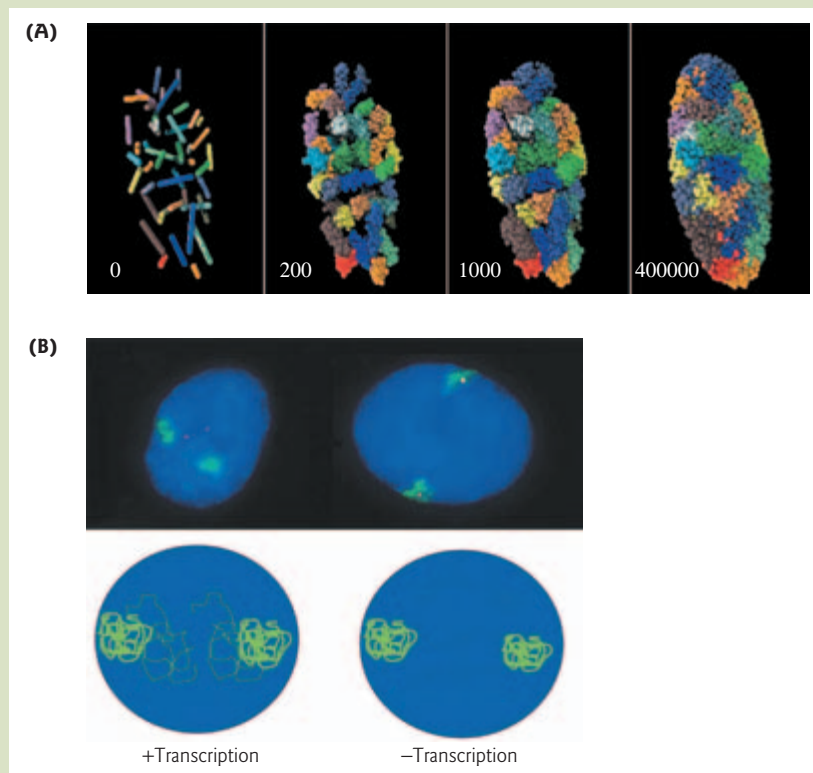
Whether the matrix is viewed as a stable, rigid, scaffold or a dynamic, transient, transcription-dependent structure depends to a large extent on the method applied for its characterization. The biological reality of the nuclear matrix remains in question. An alternative model suggests that nothing contributes as much to nuclear structure as does the chromatin itself. In other words, the “matrix” may be established by particular nuclear functions, as opposed to being present as a structural framework which then promotes function.





The chromatin fiber in a typical human chromosome is long enough to pass many times around the nucleus, even when condensed into loops (see Fig. 3.2). Chromosome “painting” – *in situ* hybridization with chromosome-specific probes – has shown that in the nucleus, each chromosome occupies its own distinct region or “territory” (Fig. 1). The territories do not generally intermingle. Specific genes do not always occupy the same relative position in three-dimensional space; however, in many vertebrates, chromosomes with low gene density reside at the nuclear periphery, whereas chromosomes with high gene density are located in the nuclear interior. Transcription appears to drive

decondensation of chromatin territories. One model proposes that the DNA loops that form in the decondensed regions are associated with transcription “factories,” containing a number of actively transcribed genes, RNA polymerase, and associated factors. It is estimated that ~16 loops would be associated with each factory. The loops often seem to surround the factory in a “cloud.” These factories are associated with the underlying nuclear matrix (see Focus box 11.2). Transcriptionally active genes also appear to be preferentially associated with nuclear pore complexes (see Focus box 11.6). This may promote direct entry of pre-mRNAs into the processing and nuclear export pathways.



**Figure 1 Chromosome territories and transcription factories.** (A) Simulation of a human model nucleus based on results from 24-color chromosome fluorescent *in situ* hybridization (FISH). The first image shows 46 statistically placed rods representing the 46 human chromosomes. The next images simulate the decondensation process and show the resulting chromosome territory arrangement. (Reprinted from Bolzer, A., Kreth, G., Solovei, I. et al. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology* 3:826–842.) (B) Transcription decondenses chromatin. (Upper panel) Loci (red) in areas of very high transcriptional activity are frequently found outside chromosome territories (green). When transcription is blocked (e.g. with an inhibitor), loci are now found more frequently within chromosome territories. (Lower panels) A model is shown suggesting that transcription decondenses chromosome territories, extruding large chromatin loop domains in a “cloud.” These loops may associate with transcription “factories” containing a number of actively transcribed genes. The loops collapse back into condensed territories when transcription ceases. (Reprinted from Chubb, J.R., Bickmore, W.A. 2003. Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* 112:403–406. Copyright © 2003, with permission from Elsevier.)



- 2** Transcription factors (see Section 11.5): sequence-specific DNA-binding proteins that bind to gene promoters and long-range regulatory elements and mediate gene-specific transcriptional activation or repression.
- 3** Transcriptional coactivators and corepressors (see Section 11.6): proteins that increase or decrease transcriptional activity through protein–protein interactions without binding DNA directly. Coactivators are operationally defined as components required for activator-directed (“activated”) transcription, but dispensable for activator-independent (“basal”) transcription. Coactivators and corepressors either serve as scaffolds for the recruitment of proteins containing enzymatic activities, or they have enzymatic activities themselves for altering chromatin structure. They include chromatin remodeling and modification complexes that assist the transcriptional apparatus to bind and move through chromatin.

## Components of the general transcription machinery

RNA polymerase II (RNA pol II) is recruited to specific promoters at the right time by extremely elaborate machineries to catalyze RNA synthesis. The polymerase and a host of other factors, including the general transcription factors, work together to form a preinitiation complex on core promoters and to allow subsequent transcription initiation (Fig. 11.8). Over the last 20 years the three major components of the general (basal) transcription machinery have been identified by techniques such as biochemical fractionation of cell extracts and *in vitro* transcription assays:

- 1** RNA pol II: a 12-subunit polymerase capable of synthesizing RNA and proofreading nascent transcript.
- 2** General transcription factors: a set of five general transcription factors, denoted TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH, is responsible for promoter recognition and for unwinding the promoter DNA. The nomenclature denotes “transcription factor for RNA polymerase II,” with letters designating the individual factors. RNA pol II is absolutely dependent on these auxiliary transcription factors for the initiation of transcription.
- 3** Mediator: a 20-subunit complex, which transduces regulatory information from activator and repressor proteins to RNA pol II.

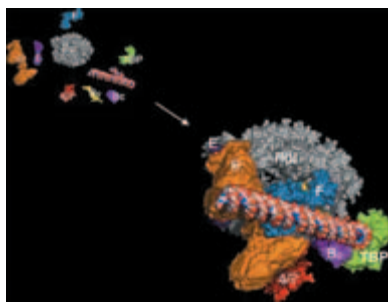
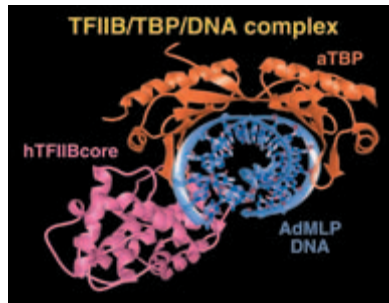
## Structure of RNA polymerase II

Among the three nuclear eukaryotic RNA polymerases, RNA pol II is the best characterized. The most detailed analysis has been completed for RNA pol II from the budding yeast *Saccharomyces cerevisiae*. The 0.5 MDa enzyme complex consists of 12 subunits (Rpb1 to 12), numbered according to size, that are highly conserved among eukaryotes. Crystal structures have revealed that yeast RNA pol II has two distinct structures and can be dissociated into a 10-subunit catalytic core and a heterodimer of subunits Rpb4 and Rpb7 (Rpb4/7 complex) (Fig. 11.9). The RNA pol II core enzyme is catalytically active but requires the Rpb4/7 complex and the general transcription factors for initiation from promoter DNA. Rpb4/7 functions at the interface of the transcriptional and post-transcriptional machinery, playing a part in mRNA nuclear export and transcription-coupled DNA repair (see Section 7.6). An additional component of RNA pol II, the mobile C-terminal domain (CTD) of Rpb1, is not seen in crystals because it is unstructured.

### RNA polymerase II catalytic core

The three-dimensional structure of the 10-subunit core enzyme of yeast RNA pol II was reported in 2001. The structure has been determined both alone and with DNA and RNA, in the form of a transcribing complex (Fig. 11.9). The two large subunits Rpb1 and Rpb2 form the central mass of the enzyme and a positively charged “cleft.” The nucleic acids occupy this deep cleft, with 9 bp of RNA–DNA hybrid at the center. One side of the cleft is formed by a massive, mobile protein element termed the “clamp.” The active center is formed between the clamp, a “bridge helix” that spans the cleft, and a “wall” of protein density

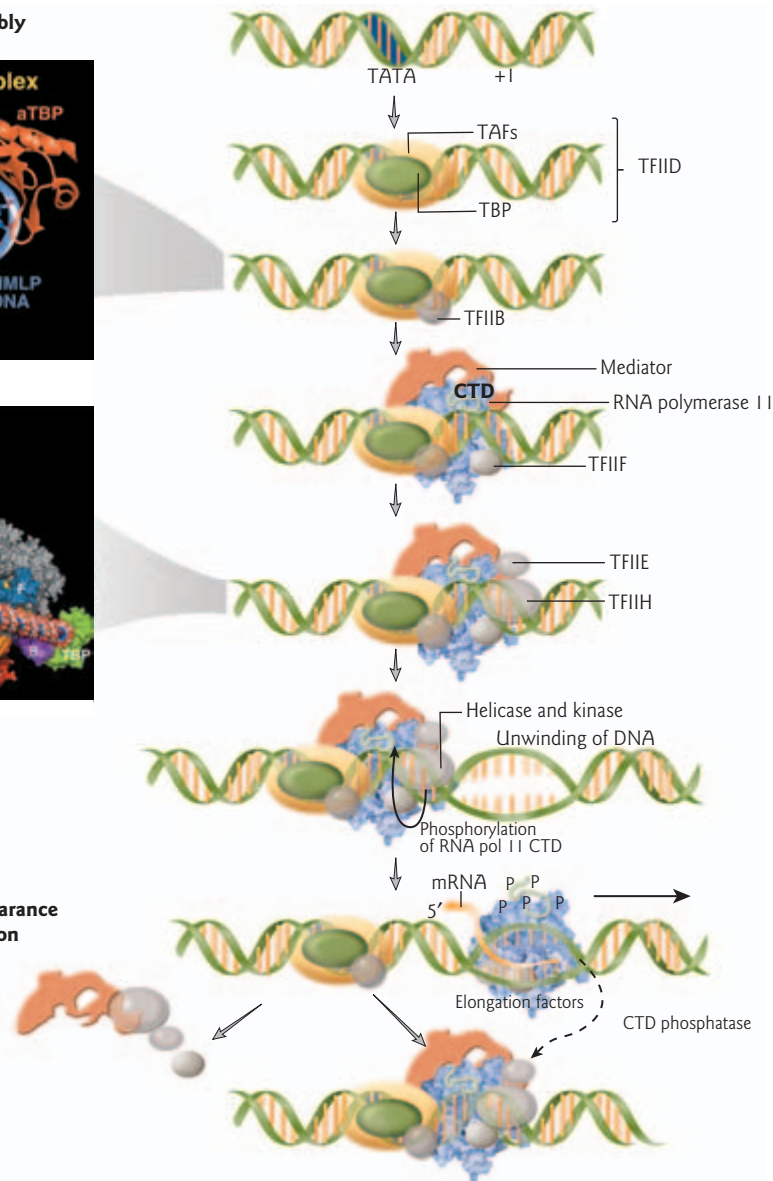
## ① Preinitiation complex assembly



## ② Initiation

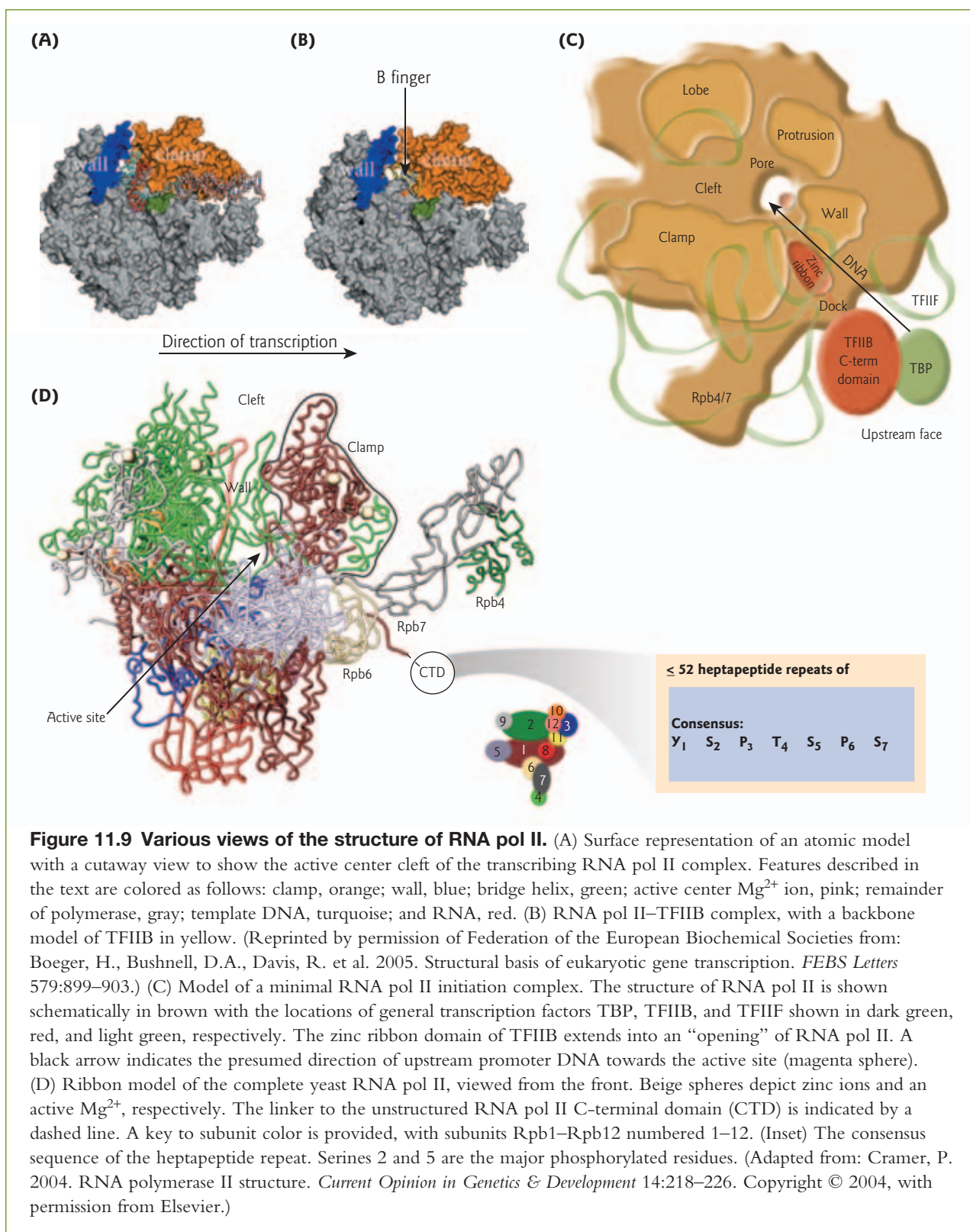
## ③ Promoter clearance and elongation

## ④ Reinitiation



**Figure 11.8 Preinitiation complex formation and initiation of transcription.** (1) Assembly of a stable preinitiation complex for RNA pol II transcription. Binding of TFIID to the promoter provides a platform to recruit TFIIB, TFIIF together with RNA pol II (in a complex with Mediator), and then TFIIE and TFIIH. (Upper inset) The saddle-like structure of the TATA-binding protein (TBP) bound to a TATA-containing sequence in DNA, which it unwinds and bends sharply. TAF, TBP-associated factor. (Image courtesy of Song Tan, Pennsylvania State University). (Lower inset) A “minimal” initiation complex of RNA pol II and the general transcription factors from combined results of X-ray diffraction and electron microscopy. (Reprinted by permission of Federation of the European Biochemical Societies from: Boeger, H., Bushnell, D.A., Davis, R. et al. 2005. Structural basis of eukaryotic gene transcription. *FEBS Letters* 579:899–903.) (2) Initiation. The helicase activity of TFIIH unwinds the DNA allowing its transcription into RNA, and its kinase activity phosphorylates the C-terminal domain (CTD) of RNA pol II. (3) Promoter clearance and elongation. As the polymerase moves away from the promoter to transcribe the gene, TFIID remains bound at the TATA box allowing the formation of a new stable complex and further rounds of transcription. (4) Reinitiation requires dephosphorylation of the RNA pol II CTD.





that blocks the end of the cleft. Because the “wall” prevents straight passage of nucleic acids through the cleft, the axis of the RNA–DNA hybrid is at nearly 90° to that of the entering DNA duplex. A “pore” beneath the active site widens towards the outside, like an inverted funnel. The rim of the pore includes a loop of Rpb1 that binds a metal ion ( $Mg^{2+}$ ). A second metal ion can bind weakly further in the pore. A linker that connects to the CTD of Rpb1 extends from the base of the clamp. Rpb4/7 is located on the core enzyme surface, below the clamp.

### RNA polymerase II C-terminal domain (CTD)

The CTD is a unique tail-like feature of the largest RNA pol II subunit, which consists of up to 52 heptapeptide repeats of the amino acid consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (see Fig. 11.9 inset). The CTD is required for mRNA processing *in vivo* and has been shown to bind to processing factors *in vitro* (see Fig. 13.6). During the transcription cycle, the CTD undergoes dynamic phosphorylation of serine residues at positions 2 and 5 in the repeat (see Fig. 11.8). Phosphorylation is catalyzed by TFIIF and other kinases, and dephosphorylation is catalyzed by the phosphatase Fcp1. Transcription initiation requires an unphosphorylated CTD, whereas elongation requires a phosphorylated CTD. When phosphorylated, the CTD can bind mRNA processing factors during termination events (see Section 13.5). For recycling of RNA pol II and reinitiation of transcription, the CTD must again be dephosphorylated (see Fig. 11.8).

### General transcription factors and preinitiation complex formation

General transcription factors have been defined biochemically as factors required for the correct initiation of RNA pol II transcription *in vitro* on a promoter with a classic TATA box and a strong initiator (Inr) element. Crystallographic studies suggest that small domains of the general transcription factors can enter “openings” in RNA pol II to modulate its function during transcription initiation. Assembling the general transcription apparatus involves a series of highly ordered steps (see Fig. 11.8). Binding of TFIID provides a platform to recruit other general transcription factors and RNA pol II to the promoter. *In vitro*, these proteins assemble at the promoter in the following order: TFIIB, TFIIF together with RNA pol II (in a complex with yet more proteins such as Mediator), and then TFIIE and TFIIF which bind downstream of RNA pol II. In some cases, TFIIA is recruited prior to TFIIB and contributes to complex stability. Because TFIIA (and its subunit TFIIF) is not absolutely required for preinitiation complex formation and transcription initiation *in vitro*, it is not typically considered a general transcription factor. TFIID and the other general transcription factors are not sufficient to reconstitute DPE-dependent transcription. Two additional factors, a protein kinase (CK2) and a coactivator protein (PC4), are necessary to initiate DPE-dependent transcription.

### TFIID recruits the rest of the transcriptional machinery

The first general transcription factor to associate with template DNA is TFIID (see Fig. 11.8). TFIID is a complex composed of the TATA-binding protein (TBP) and 14 TBP-associated factors (TAFs). TBP is a sequence-specific DNA-binding protein (see Fig. 9.15B,D) that recognizes the TATA box, as well as some other core promoter elements. TBP is highly conserved through evolution and appears to be required for transcription by all three eukaryotic RNA polymerases. TBP contains an antiparallel  $\beta$ -sheet that sits on the DNA like a saddle in the minor groove and bends the DNA (see Fig. 11.8). The most common recognition motif between a transcription factor and DNA is the  $\alpha$ -helix domain of the protein and base pairs within the major groove (see Section 11.5). TBP is a notable exception to this general rule. The seat of TBP’s “saddle” is the site of many protein–protein interactions both with the TAFs and the general transcription factors.

Binding of TFIID to the core promoter is a critical rate limiting step at which activators and/or chromatin remodeling factors can control transcription (see section on “Mediator” below). Transcription initiation can also be regulated at the level of preinitiation complex formation at core promoters.



Multicellular eukaryotes have evolved multiple related TFIID complexes that can function at distinct promoters through the use of several distinct TBP-like factors and tissue-specific TAFs. These TAFs interact with the transactivation domains of specific transcription factors (see Section 11.5). For example, the glutamine-rich transactivation domain of Sp1 binds to TAF<sub>II</sub>110 in the old nomenclature).

### TFIIB orients the complex on the promoter

Upon TFIID binding, TFIIB joins the growing assemblage (see Fig. 11.8). The TATA box promoter element to which TBP binds has a rough two-fold rotational symmetry (see Table 11.2). TBP could in principle align with it in either of two directions on either strand of the DNA double helix. Instead, binding of TFIIB orients the complex via specific protein–protein and protein–DNA contacts. TFIIB binds to only one end of TBP and simultaneously to a GC-rich DNA sequence that follows the TATA motif. The TFIIB–TBP–DNA complex signposts the direction for the start of transcription by RNA pol II and indicates which strand of the DNA double helix acts as the template (see Fig. 11.8). The N-terminal “zinc ribbon” domain (a cysteine-rich, zinc-binding region) of TFIIB binds RNA pol II and is essential for its recruitment (see Fig. 11.9). This brings the initiation complex to a point on the surface of RNA pol II from which the DNA need only follow a straight path to the active site. Because of the conserved spacing from the TATA box to the transcription start site (+1), the start site is positioned in the polymerase active center.

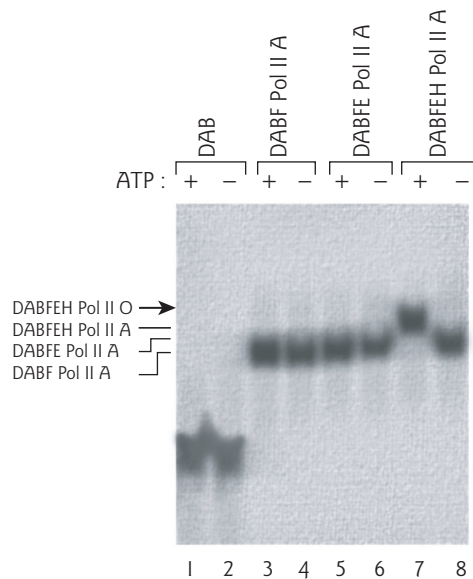
### TFIIE, TFIIIF, and TFIIH binding completes the preinitiation complex formation

Promoter loading onto RNA pol II also requires TFIIE, which forms a tight complex with the polymerase (see Fig. 11.8). Entry of TFIIB, core promoter DNA, and TFIIIF in the growing assemblage leads to binding of TFIIE which, in turn, recruits TFIIH. TFIIH is the most complex of all the general transcription factors. It contains at least nine polypeptide subunits with diverse functions. These include a cyclin-dependent protein kinase (CDK7), cyclin H, a 5′ → 3′ helicase, a 3′ → 5′ helicase, two zinc finger proteins, and a ring finger protein. Danny Reinberg and co-workers demonstrated that TFIIH is the protein kinase that phosphorylates the CTD of RNA pol II, by showing that purified TFIIH could convert the unphosphorylated polymerase to its phosphorylation form *in vitro* (Fig. 11.10). ATP-dependent helicase activity was demonstrated for yeast TFIIH (encoded by the *RAD25* gene) by Satya Prakash and colleagues. They overexpressed the protein in yeast, purified it almost to homogeneity, and then showed that the protein had helicase activity *in vitro* (Fig. 11.11).

TFIIH functions not only in transcription but also in DNA repair (see Fig. 7.11). Upon binding of these factors, the complex of core promoter, general transcription factors, and RNA pol II (in association with Mediator, see below) is called the preinitiation complex (PIC). Promoter melting (unwinding of the double-stranded DNA) requires hydrolysis of ATP. This reaction is mediated by the ATPase/helicase subunit of TFIIH, with the help of TFIIE. Unwinding is followed by “capture” of the nontemplate strand by TFIIIF. The template strand descends to the active site of RNA pol II (see Fig. 11.9).

### Mediator: a molecular bridge

*In vitro* transcription assays can be performed using a minimum set of general transcription factors and purified core RNA pol II. In such assays, RNA pol II and associated factors can stimulate low levels of transcription (referred to as basal transcription). However, under these conditions, the core RNA pol II is not responsive to transcriptional activators that can increase the frequency of initiation or can increase gene activity *in vivo*. Following up on these observations, Roger Kornberg and colleagues provided the first evidence that an additional factor is required for activator-responsive transcription (Fig. 11.12). This led to the characterization of a protein complex called “Mediator” that serves as a molecular bridge that connects transcriptional activators bound at enhancers, or other long-range regulatory elements, with RNA pol II (Fig. 11.13). Mediator is not required for basal RNA pol II transcription from a core promoter *in vitro* and



### Figure 11.10 RNA pol II

**phosphorylation.** Electrophoretic mobility shift assays were performed with a  $^{32}\text{P}$ -end-labeled DNA fragment containing a viral promoter, and preinitiation complexes containing the unphosphorylated form of RNA polymerase II (Pol IIA) and various combinations of TFIID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIH (DABFEH) in the presence and absence of ATP as indicated at the top (see Fig. 9.15A for methods). Only when TFIH was present did ATP shift the mobility of the complex (compares lane 7 and 8). The simplest explanation is that TFIH promotes phosphorylation of the input polymerase from its unphosphorylated form to its phosphorylated form (Pol IIO). Additional experiments by Reinberg and colleagues demonstrated directly that TFIH phosphorylates RNA pol II. (Reprinted by permission from Nature Publishing Group and Macmillan Publishers Ltd: Lu, H., Zawel, L., Fisher, L., Egly, J.M., Reinberg, D. 1992. Human general transcription factor IIF phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358:641–645. Copyright © 1992.)

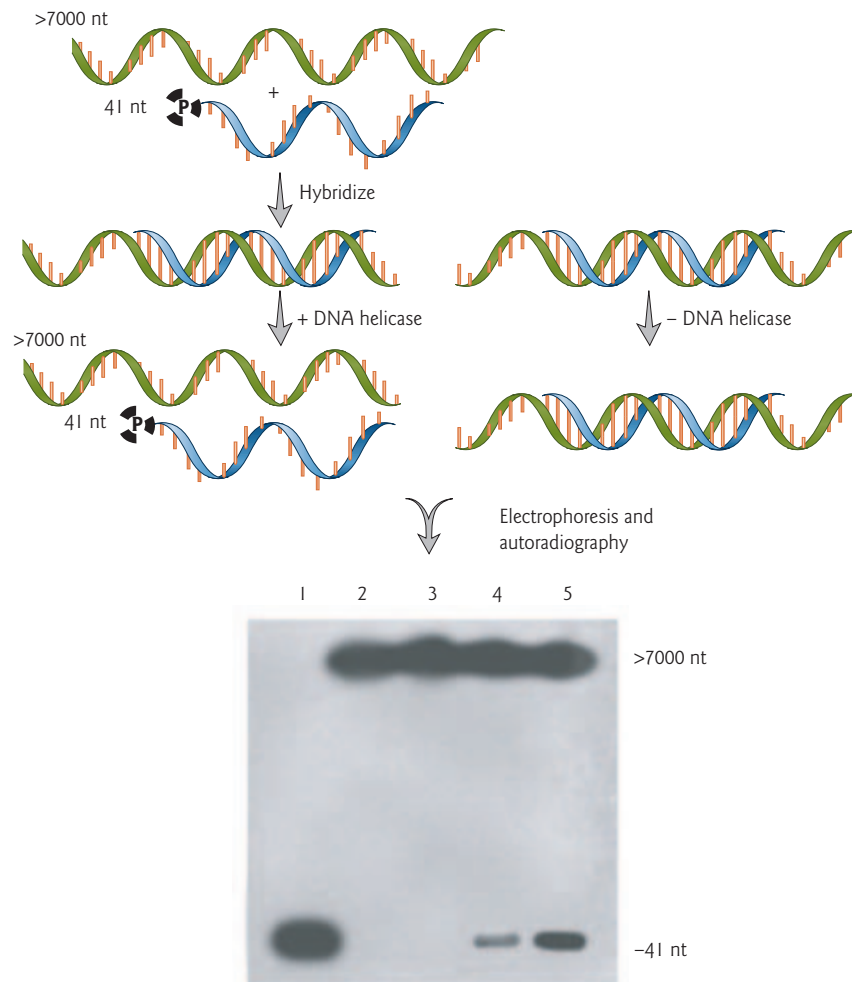
hence is not classified as a general transcription factor. It is often referred to as a coactivator; however, this classification is a misnomer because Mediator is a component of the preinitiation complex and several Mediator subunits are required for transcription of almost all genes. In addition to promoting preinitiation complex assembly, Mediator also stimulates the kinase activity of TFIH.

RNA polymerase II and the general transcription factors have counterparts in bacteria (see Section 10.3); however, Mediator is unique to eukaryotes. Mediator is expressed ubiquitously in eukaryotes from yeast to mammals. While yeast has one such complex, multicellular eukaryotes contain several related complexes. At least seven different mammalian Mediator complexes consisting of 25–30 proteins similar to yeast Mediator were subsequently identified in many labs. They include previously characterized complexes such as TRAP–SMCC (thyroid hormone receptor-associated protein/SRB–Med-containing cofactor) and DRIP (vitamin D receptor-interacting proteins) complexes. A unified nomenclature has been proposed, renaming these complexes as Mediator T/S and Mediator D, respectively, to more clearly designate their function in the cell. The majority of Mediator complexes act as transcriptional coactivators (e.g. Mediator D). However, there is a Mediator complex that represses transcription and Mediator T/S can act as both a repressor and an activator depending on the conditions used in the cell-free transcription assay.

## 11.5 Transcription factors

The regulation of gene activity at the transcriptional level generally occurs via changes in the amounts or activities of transcription factors. Of course, the genes encoding the transcription factors themselves may be transcriptionally induced or repressed by other regulatory proteins, or the transcription factors may be activated or deactivated by proteolysis, covalent modification, or ligand binding. Transcription factors



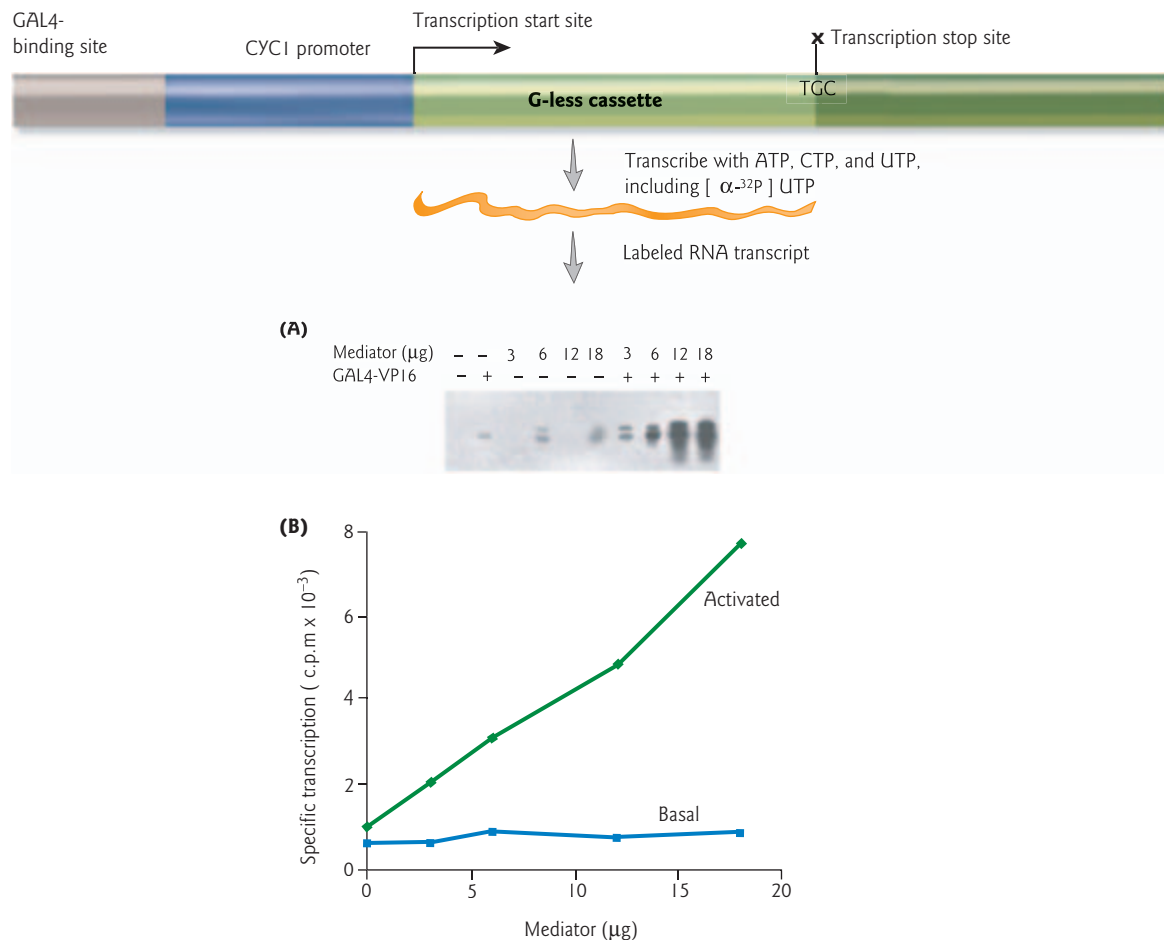


**Figure 11.11 TFIID helicase activity.** A labeled 41 nt piece of single-stranded DNA (blue) was hybridized to its complementary region in a much larger (> 7000 nt), unlabeled M13 phage DNA. This substrate was then incubated with or without RAD25 protein (the yeast TFIID) in the presence and absence of ATP. DNA helicase activity unwinds the partial duplex and releases the labeled 41 nt DNA from the phage DNA. The 41 nt DNA is distinguished from the hybrid (> 7000 nt) by gel electrophoresis. The results of the helicase assay are shown: lane 1, heat-denatured 41 nt DNA (size marker); lane 2, no protein (negative control); lane 3, 20 ng RAD25 with no ATP; lane 4, 10 ng of RAD25 plus ATP; lane 5, 20 ng of RAD25 plus ATP. (Reprinted by permission from Nature Publishing Group and Macmillan Publishers Ltd: Guzder, S.N., Sung, P., Bailly, V., Prakash, L., Prakash, S. 1994. RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369:578–581. Copyright © 1994.)

influence the rate of transcription of specific genes either positively or negatively (activators or repressors, respectively) by specific interactions with DNA regulatory elements (see Section 11.3) and by their interaction with other proteins.

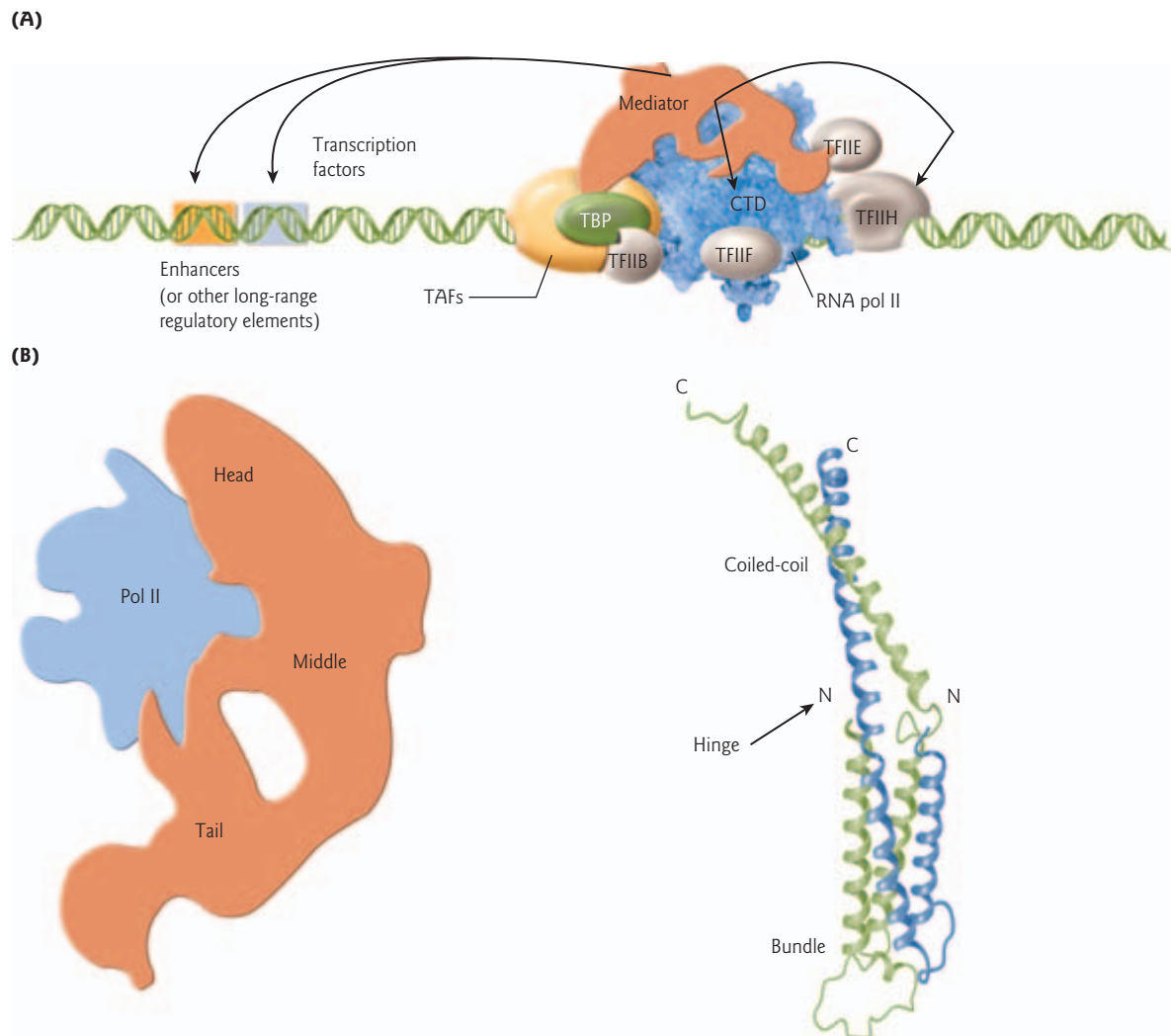
### Transcription factors mediate gene-specific transcriptional activation or repression

Transcription factors that serve as repressors block the general transcription machinery, whereas transcription factors that serve as activators increase the rate of transcription by several mechanisms:



**Figure 11.12 The discovery of Mediator.** The yeast *CYC1* promoter was placed downstream of a GAL4-binding site and upstream of a G-less cassette, so transcription of the G-less cassette depended on both the *CYC1* promoter and GAL4. The G-less cassette assay is a variation of a run-off transcription assay (see Fig. 11.3). Instead of cutting the template DNA with a restriction endonuclease, a stretch of nucleotides lacking guanine in the nontemplate strand is inserted downstream of the promoter. This template is transcribed *in vitro* with CTP, ATP, and UTP, one of which is labeled, but no GTP. Transcription will stop at the end of the cassette where the first G is required, yielding a transcript of predictable size (based on the size of the G-less cassette, which is usually a few hundred base pairs long) on a polyacrylamide gel. The more transcript produced, the stronger will be the corresponding band on the autoradiograph. In this experiment, the construct was transcribed in the presence of a Mediator-containing fraction from yeast cells in the amounts indicated, and in the absence (–) or presence (+) of the activator protein GAL4-VP16 as indicated. (A) Phosphorimager scan of the gel. (B) Graphic presentation of the results in (A). Mediator greatly stimulates transcription in the presence of the activator (lanes 7–10), but has no effect on unactivated (basal) transcription (lanes 3–6). (Reprinted by permission from Nature Publishing Group and Macmillan Publishers Ltd: Flanagan, P.M., Kelleher, R.J. III, Sayre, M.H., Tschochner, H., Kornberg, R.D. 1991. A mediator required for activation of RNA polymerase II transcription *in vitro*. *Nature* 350:436–438. Copyright © 1991.)

- 1 Stimulation of the recruitment and binding of general transcription factors and RNA pol II to the core promoter to form a preinitiation complex.
- 2 Induction of a conformational change or post-translational modification (such as phosphorylation) that stimulates the enzymatic activity of the general transcription machinery.



**Figure 11.13 Mediator: a molecular bridge.** (A) Mediator serves as a molecular bridge between the transactivation domains of various transcription factors and RNA pol II. (B) Structure of the Mediator complex. (Left panel) Outline of Mediator bound to RNA pol II according to electron microscopy. The suggested head, middle, and tail modules are indicated (Redrawn from Baumli, S., Hoepfner, S., Cramer, P. 2005. A conserved Mediator hinge revealed in the structure of the MED7/MED21 (Med7/Srb7) heterodimer. *Journal of Biological Chemistry* 280:18171–18178.) (Right panel) A ribbon model of the MED7 (blue)/MED21 (green) heterodimer is shown, based on the 3.0 Å crystal structure. The heterodimer structure has a very extended four helix-bundle domain and a coiled-coil protrusion, connected by a flexible hinge. It spans one-third the length of Mediator and almost the diameter of RNA pol II. The hinge may account for changes in Mediator structure upon binding RNA pol II or transcription factors (Protein Data Bank, PDB:1YKH).

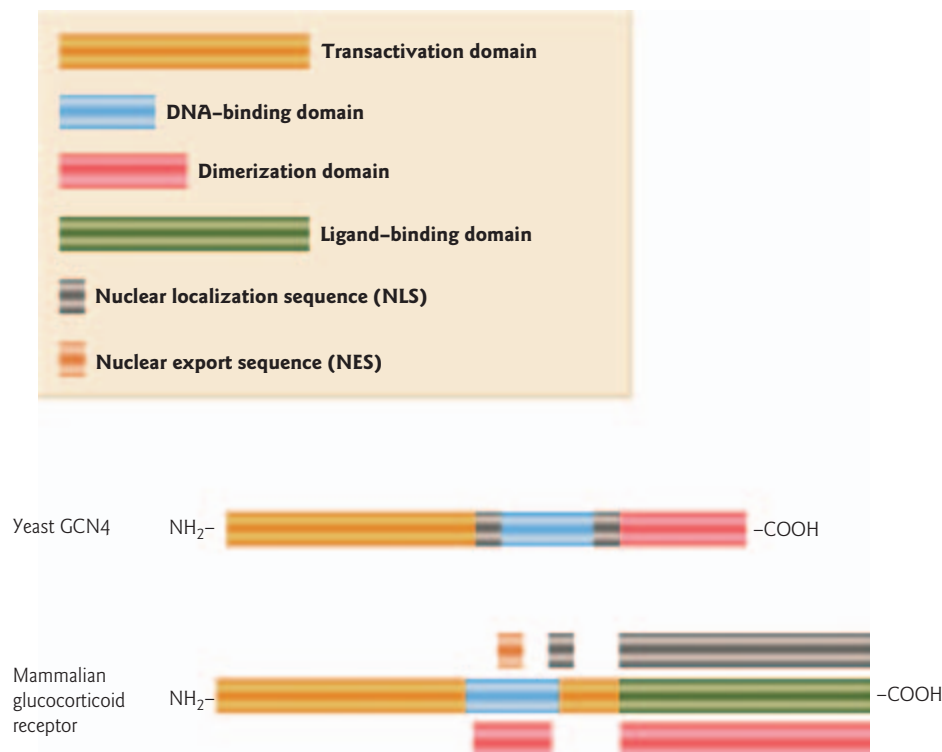
### 3 Interaction with chromatin remodeling and modification complexes to permit enhanced accessibility of the template DNA to general transcription factors or specific activators.

These different roles can be promoted directly via protein–protein interaction with the general transcription machinery (see Section 11.4) or via interactions with transcriptional coactivators and corepressors (see Section 11.6).

Many transcription factors are members of multiprotein families. For example, nuclear receptors are members of a superfamily of related proteins, including the receptors for steroid hormones, thyroid hormone, and vitamin D. NF- $\kappa$ B is yet another family of proteins (see Section 11.10), and Sp1 – one of the first transcription factors to be isolated – is a member of the Sp family of proteins. Within each family, the members often display closely related or essentially identical DNA-binding properties but distinct activator or repressor properties.

## Transcription factors are modular proteins

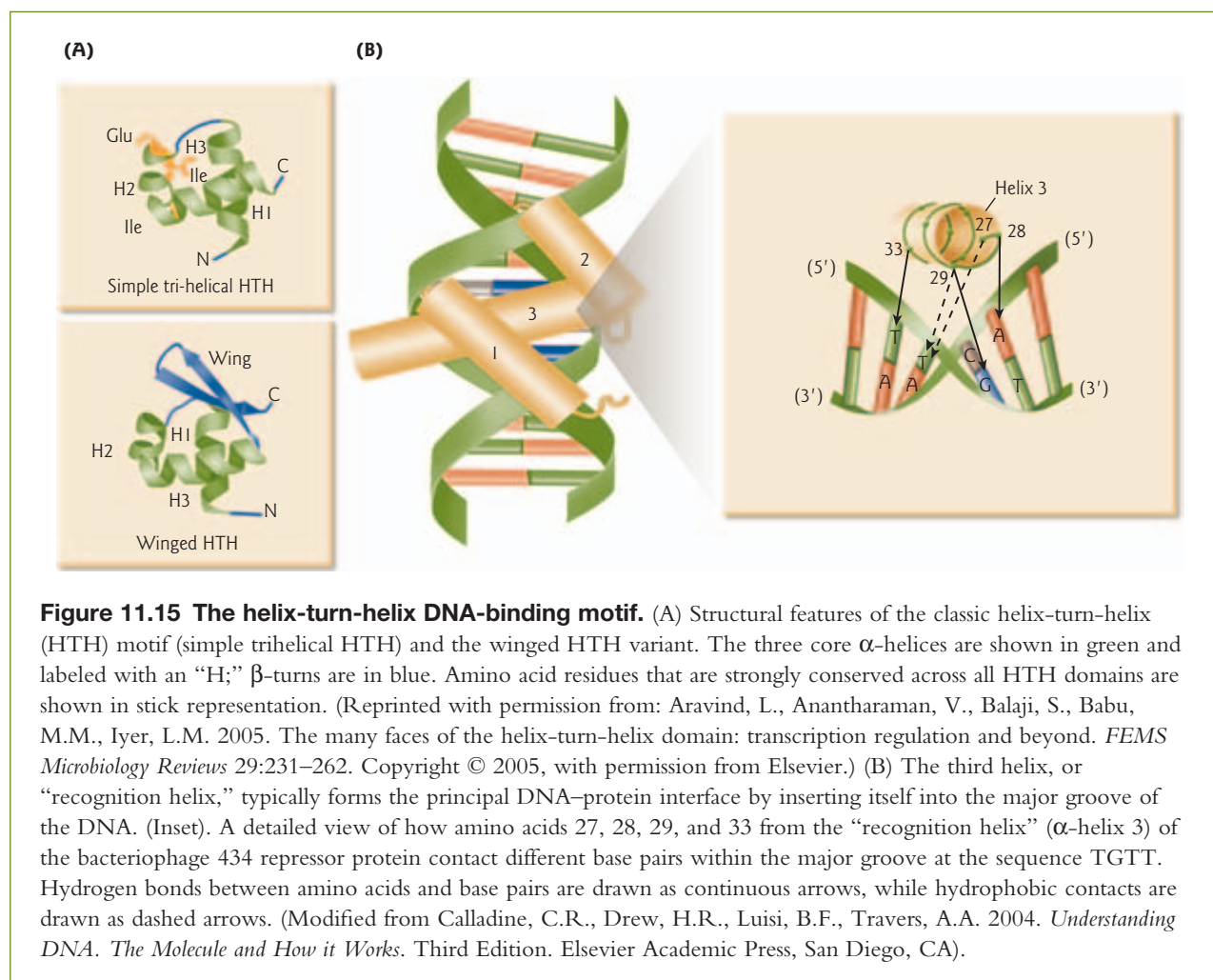
Transcription factors are modular proteins consisting of a number of domains (Fig. 11.14). Recognition of this feature triggered the development of a powerful technique for analyzing protein–protein interactions *in vivo* – the yeast two-hybrid assay (see Fig. 9.16B). The three major domains are a DNA-binding domain, a transactivation domain, and a dimerization domain. In addition, transcription factors typically have a nuclear localization sequence (NLS), and some also have a nuclear export sequence (NES) (see Section 11.9). Some transcription factors also have ligand-binding (regulatory) domains, such as hormone-binding domains, which are essential for controlling their activity.



**Figure 11.14 Transcription factors are composed of modular domains.** Transcription factors are composed of separable, functional components and typically include a DNA-binding domain, a transactivation domain, and a dimerization domain. Transcription factors may contain more than one transactivation domain, but rarely contain more than one DNA-binding domain. In addition, transcription factors usually have a nuclear localization sequence (NLS), and some also have a nuclear export sequence (NES) (see Section 11.10). Some transcription factors also have ligand-binding (regulatory) domains, such as hormone-binding domains, which are essential for controlling their activity. The domain structure of the yeast GCN4 transcription factor and the mammalian glucocorticoid receptor are compared. Both GCN4 and the glucocorticoid receptor have two NLS motifs. Functional domains may overlap, as shown by stacked colored boxes for the glucocorticoid receptor.

## DNA-binding domain motifs

The DNA-binding domain positions a transcription factor on a specific DNA sequence. Hundreds of protein–DNA complexes have now been analyzed by X-ray crystallography. In addition, NMR spectroscopy has been used to study complexes in solution (see Section 9.10 for methods). These studies have provided a detailed picture of how the DNA-binding domain interacts specifically with the bases of DNA. High-affinity binding is dependent on the overall three-dimensional shape and formation of specific hydrogen bonds. The amino acids of a protein can make specific hydrogen bonds with exposed atoms on the sides of the base pairs or along the “floor” of the major or minor groove in the DNA. The most common recognition pattern between transcription factors and DNA is an interaction between an  $\alpha$ -helical domain of the protein and about five base pairs within the major groove of the DNA double helix (Fig. 11.15). The  $\alpha$ -helical domain is complementary in its shape to the surface of the DNA formed by base pairs and phosphates. For high-affinity binding, both surfaces must match closely in terms of hydrogen bonds and hydrophobic contacts. Before the transcription factor and DNA come together, their “polar groups” (e.g. N-H, O-H, N or O) form hydrogen bonds to surrounding water molecules. The hydrogen bonds to water are mostly replaced in the transcription factor–DNA complex by hydrogen bonds made directly between protein and DNA. The replacement of these water molecules lends stability to the complex. Loss of just a few hydrogen bonds or hydrophobic contacts from a specific transcription factor–DNA complex will usually result in a large loss of specificity for that particular DNA regulatory element.



When the DNA-binding domains are compared, many transcription factors fall into groups defined by related “motifs.” A motif is defined as a cluster of amino acid residues that has a characteristic three-dimensional folding pattern and carries out a specific function. The following is an overview of some of the more common, well described, sequence-specific DNA-binding motifs: the helix-turn-helix, zinc finger, basic leucine zipper, and basic helix-loop-helix motifs.

### Helix-turn-helix (HTH)

The HTH motif was the first DNA-binding domain to be well characterized. It was initially identified in 1982 by comparison of the structures of CAP (*E. coli* catabolite activator protein) and Cro (repressor protein from bacteriophage  $\lambda$ ). The HTH motif is now known to be present in most prokaryotic regulatory proteins, including the *E. coli* Lac repressor protein and CAP (see Figs 10.15 and 10.16). The classic HTH domain is a simple amino acid fold composed of three core  $\alpha$ -helices that form a right-handed helical bundle with a partly open configuration (Fig. 11.15). The third helix, or “recognition helix,” typically forms the principal DNA–protein interface by inserting itself into the major groove of the DNA. A characteristic sharp turn of several amino acids, called the  $\beta$ -turn, separates the second and third helix.

The homeodomain is a variant of the classic HTH that is present in many transcription factors that regulate development. It is a conserved 60 amino acid domain that is encoded by a 180 bp “homeobox” sequence in the DNA (Focus box 11.4). Other variant forms of the trihelical HTH may contain additional elaborations, such as the winged HTH motif, which was discovered in 1993. This variant is distinguished by the presence of the “wing,” which is a C-terminal  $\beta$ -strand hairpin that folds against the shallow cleft of the partially open trihelical core (Fig. 11.15A). At least 80 genes with this motif are known, many with developmentally specific patterns of expression. The winged HTH is also termed a fork head domain after the founding member of this group, the *Drosophila fork head* gene. *Fork head* mutations cause homeotic transformation of portions of the gut; e.g. the foregut and hindgut are replaced by head structures of the fruitfly. A homeotic mutation is a mutation that transforms one part of the body into another part (Focus box 11.4).

### Zinc finger (Zif)

The zinc finger structural motif is one of the most prevalent DNA-binding motifs. It was first described in 1985 for *Xenopus laevis* TFIIA – a transcription factor essential for 5S ribosomal RNA (rRNA) gene transcription by RNA polymerase III. TFIIA binds to the 5S rRNA gene promoter, as well as to the 5S rRNA itself. The name zinc finger was coined because the two-dimensional diagram of its structure resembles a finger (Fig. 11.16). A “finger” is formed by interspersed cysteines (Cys) and/or histidines (His) that covalently bind a central zinc ( $\text{Zn}^{2+}$ ) ion, folding a short length of the amino acid chain into a compact loop domain. When the three-dimensional structure was solved, it was shown that the left side of the finger folds back on itself to form a  $\beta$ -sheet. The right side twists into an  $\alpha$ -helix. Binding of zinc by cysteines in the  $\beta$ -sheet and histidines in the  $\alpha$ -helix draws the halves together near the base of the finger. It also brings hydrophobic amino acids close to one another at the fingertip where their mutual attraction helps to stabilize the motif. The finger inserts its  $\alpha$ -helical portion into the major groove of the DNA. Generally, there is a linker region of 7–8 amino acids in between each zinc finger module. The number of fingers is variable between different zinc finger-containing transcription factors.

There are a number of different types of zinc finger motifs. For example, TFIIA has nine fingers and GLI3 (Disease box 11.2) has five of the classic  $\text{Cys}_2\text{-His}_2$  ( $\text{C}_2\text{H}_2$ ) pattern described above, while nuclear receptors have two fingers of a  $\text{Cys}_2\text{-Cys}_2$  ( $\text{C}_2\text{C}_2$ ) pattern. Figure 11.16 illustrates the zinc finger DNA-binding domain of the glucocorticoid receptor. Three to four amino acids at the base of the first finger confer specificity of binding. A dimerization domain near the base of the second finger is the region that interacts with another glucocorticoid receptor to form a homodimer. Each protein in the pair recognizes half of a two-part DNA regulatory element called a glucocorticoid response element (GRE). The binding





## FOCUS BOX 11.4

## Homeoboxes and homeodomains

In the late 1970s developmental biologists began to understand the elegant simplicity of pattern formation in the developing fruitfly embryo. Clusters of genes are turned on in sequence, determining the overall body pattern – anterior to posterior, and dorsal to ventral – and then successive gene cascades send signals telling the cells in the various segments of a developing embryo what kind of structures to make, whether it be legs, wings, or antennae. A major advance was the discovery of the homeobox genes first in *Drosophila* in the late 1970s, and later in many other organisms from humans to the plant *Arabidopsis*. The name “homeobox” is derived from the fact that mutations in some of these genes result in a homeotic transformation, or a situation in which one body or plant part is replaced by another or is duplicated. Homeobox genes encode transcription factors containing a 60 amino acid DNA-binding motif called a homeodomain, a variant form of the classic helix-turn-helix motif. Homeodomain transcription factors regulate many embryonic developmental programs, including axis formation, limb development, and organogenesis. Some homeodomain transcription factors are also expressed in adult tissues such as liver, kidney, and intestine, where they are thought to play a role in the regenerative differentiation of cells. The best known homeobox gene subclass is the *Hox* family. Disruption of a *Drosophila Hox* gene can lead to a phenotype known as *Antennapedia* in which fly legs develop in place of the antennae (Fig. 1A). In humans, a mutation in the *Hoxd13* gene results in the duplication of a digit (polydactyly) resulting in the development of six fingers.

The *Hox* genes are present in large clusters. These clusters tend to contain fewer chromosomal breakpoints between human and mouse than expected by chance, which suggests that they are being held together by natural selection. Many *Hox* genes display a colinear pattern of expression determined by their position in the cluster. Clusters of homeobox genes are regulated, in part, by the Polycomb group proteins.

#### Colinear expression of homeobox genes

The clustering of homeobox genes allows a pattern of organized, differential gene transcription in which an expression gradient is achieved either spatially, temporally, or quantitatively, depending on the location of each gene within the cluster. Spatial colinearity means the position of

a gene in a cluster correlates with its expression domain, e.g. 5' and 3' *Hox* genes are typically expressed in the posterior and anterior portions of developing embryos, respectively (Fig. 1A). In temporal colinearity, genes at one end of the cluster are turned on first and genes at the opposite end of the cluster are turned on last. Quantitative colinearity refers to a situation where the first gene in a cluster displays the maximum level of mRNA expression and downstream genes exhibit progressively lower expression. Colinear expression of *Hox* genes was recently reported to be regulated by global enhancer elements that establish a long-range chromosomal domain.

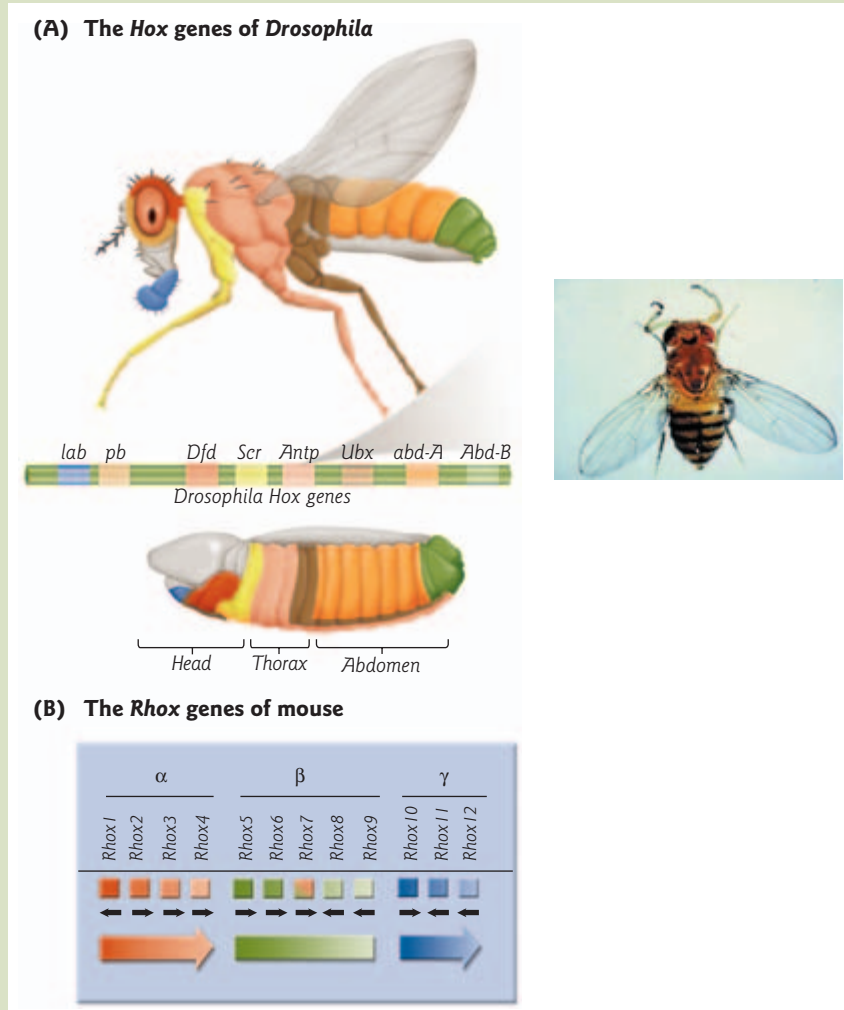
Almost 20 years after the initial discovery of the homeobox DNA element, a new cluster of 12 homeobox genes has been reported (Fig. 1B). The cluster is on the X chromosome and is a subfamily of the reproductive homeobox X-linked (*Rhox*) genes, which appear to play an important regulatory role in reproduction. They are selectively expressed in the testis, epididymus, ovary, and placenta. Their expression pattern is colinear and corresponds to their chromosomal position. Two of the three subclusters show temporal colinearity and all three show quantitative colinearity.

#### Polycomb group proteins silence homeobox genes

The *Polycomb* gene was discovered nearly 60 years ago through observations of *Drosophila* mutants. Normal male fruitflies have structures called sex combs on their front legs, used for grasping females during mating. Mutant flies were observed that also had sex combs on the second and third pairs of legs, and so were named “Polycomb.” Later work suggested that the Polycomb protein normally represses expression of the bithorax gene, and prevents structures such as sex combs or wings from forming in the wrong body segments. The Polycomb group (PcG) proteins act in conjunction with DNA sequences termed “Polycomb response elements” to maintain lineage-specific “off” transcriptional states of homeotic genes. They mediate gene silencing by altering the higher order structure of chromatin. The founding member of the PcG family, Polycomb, contains a chromodomain – a 37 amino acid motif similar to a known chromatin-binding domain in a protein called heterochromatin-associated protein 1 (HP1). Recent work shows that gene silencing in the fruitfly requires two distinct complexes of various PcG proteins. PRC2

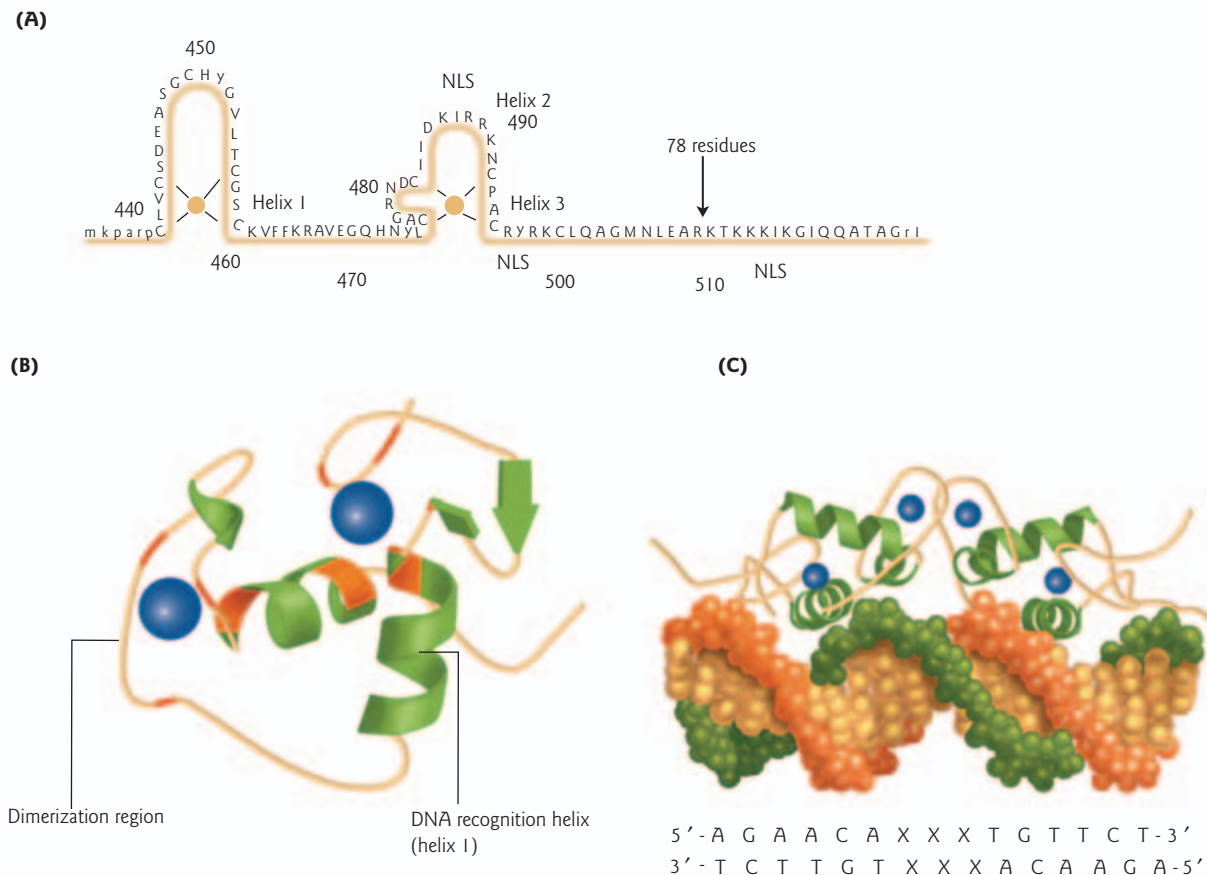
## Homeoboxes and homeodomains

## FOCUS BOX 11.4



**Figure 1 Homeobox genes.** (A) The *Hox* genes of *Drosophila*. Eight *Hox* genes regulate the identity of regions within the adult fruitfly and the embryo. The color coding represents the segments and structures that are affected by mutations in *Hox* genes (Adapted with permission by Nature Publishing Group and Macmillan Publishers Ltd: Carroll S.B. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–485. Copyright © 1995. (Inset) *Antennapedia* homeotic mutant in which the antennae are transformed into legs. (Reprinted from Carroll S.B., Grenier, J.K., Weatherbee, S.D. 2005. *From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design*. Second Edition. Blackwell Publishing, Malden, MA). (B) The *Rhox* genes of mouse. Twelve *Rhox* genes on the X chromosome are specifically expressed in male and female reproductive tissues. The genes of each subcluster ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) are expressed in the order that they occur on the chromosome during sperm differentiation, to different degrees (graded coloured arrows) and, in the  $\alpha$  and  $\gamma$  subclusters, at different times. The first genes in each subcluster are expressed earlier and to a greater maximal level than the next ones. Black arrows denote the direction of gene transcription. The pink shading for the *Rhox7* gene shows that, based on its DNA sequence, it is more closely related to the  $\alpha$ - than the  $\beta$ -cluster. (Adapted with permission by Nature Publishing Group and Macmillan Publishers Ltd: from Spitz, F., Duboule, D. 2005. Developmental biology: reproduction in clusters. *Nature* 434:715–716. Copyright © 2005.)





**Figure 11.16 The zinc finger DNA-binding domain.** (A) A schematic of the two zinc finger motifs in the DNA-binding domain of the glucocorticoid receptor (GR). The orange spheres depict the zinc ions coordinating the cysteine residues. The positions of two nuclear localization sequences (NLSs) are shown. (B) The three-dimensional structure of the GR DNA-binding domain. (Reprinted from Low, L.Y., Hernández, H., Robinson, C.V., O'Brien, R., Grossmann, J.G., Ladbury, J.E., Luisi, B. 2002. Metal-dependent folding and stability of nuclear hormone receptor DNA-binding domains. *Journal of Molecular Biology* 319:87–106. Copyright © 2002, with permission from Elsevier.) (C) The three-dimensional structure of a GR homodimer bound to DNA. (Protein Data Bank, PDB:1GDC, 1LAT). The base sequence of the glucocorticoid response element (GRE) in the DNA recognized by GR is given below. The half sites in the elements are alike if their base pairs are read along opposite strands of the DNA, in the 5' → 3' direction. One GR monomer binds to each half site.

proto-oncogene *c-jun*. Proto-oncogenes are genes that have a normal function in the cell but can become cancer-causing when they undergo certain mutations (see Section 17.2).

The bZIP motif is not itself the DNA-binding domain of the transcription factor and does not directly contact the DNA. Instead, it plays an indirect structural role in DNA binding by facilitating dimerization of two similar transcription factors. The joining of two bZIP-containing transcription factors results in the correct positioning of the two adjacent DNA-binding domains in the dimeric complex (Fig. 11.17A). The bZIP motif is a stretch of amino acids that folds into a long  $\alpha$ -helix with leucines in every seventh position. The leucines form a hydrophobic “stripe” on the face of the  $\alpha$ -helix. Two polypeptide chains with this motif coil around each other to form a dimer in a “coiled coil” arrangement. The amino acids protrude like knobs from one  $\alpha$ -helix and fit into complementary holes between the knobs on the partner helix, like a zipper. The dimer forms a Y-shaped structure and one end of each  $\alpha$ -helix protrudes into the major groove of the DNA. This allows the two basic binding regions (rich in arginine and lysine) to contact the DNA.



## DISEASE BOX 11.2

## Greig cephalopolysyndactyly syndrome and Sonic hedgehog signaling

Greig cephalopolysyndactyly syndrome is a very rare disorder that is inherited in an autosomal dominant pattern. The disorder is characterized by physical abnormalities affecting the fingers, toes, head and face. The range and severity of symptoms may vary greatly from patient to patient. Patients often have a prominent forehead (frontal bossing), an abnormally broad nasal bridge, widely spaced eyes (ocular hypertelorism), extra fingers and/or toes (polydactyly), and webbing and/or fusion of the fingers and toes (syndactyly) (Fig. 1A). This syndrome can be caused by many types of mutations in the *GLI3* gene, including translocations, large deletions, exonic deletions and duplications, small in-frame deletions, and missense, frameshift/nonsense, and splicing mutations. *GLI3* is a member of the GLI-Kruppel family of transcription factors. The *GLI1*, *GLI2*, and *GLI3* genes all encode transcription factors that share five highly conserved tandem Cys<sub>2</sub>-His<sub>2</sub> zinc fingers. The *GLI1* gene was originally identified as an amplified gene in a malignant glioblastoma, hence the acronym. *GLI3* acts as both a transcriptional activator and transcriptional repressor of downstream targets in the Sonic hedgehog pathway during development (Fig. 1B).

Embryogenesis is regulated by a number of complex signaling cascades that are critical for normal development. One such pathway begins with the secreted protein “Sonic hedgehog.” Sonic hedgehog is one of three vertebrate homologs to the *Drosophila hedgehog* gene. The original *hedgehog* gene was named for the appearance of the mutant phenotype in which a *Drosophila* embryo is covered with pointy denticles resembling a hedgehog. The first two

homologs of *hedgehog*, *desert hedgehog* and *indian hedgehog*, were named after species of hedgehog. The third homolog was named for “Sonic the hedgehog,” a character in a popular video game from Sega Genesis.

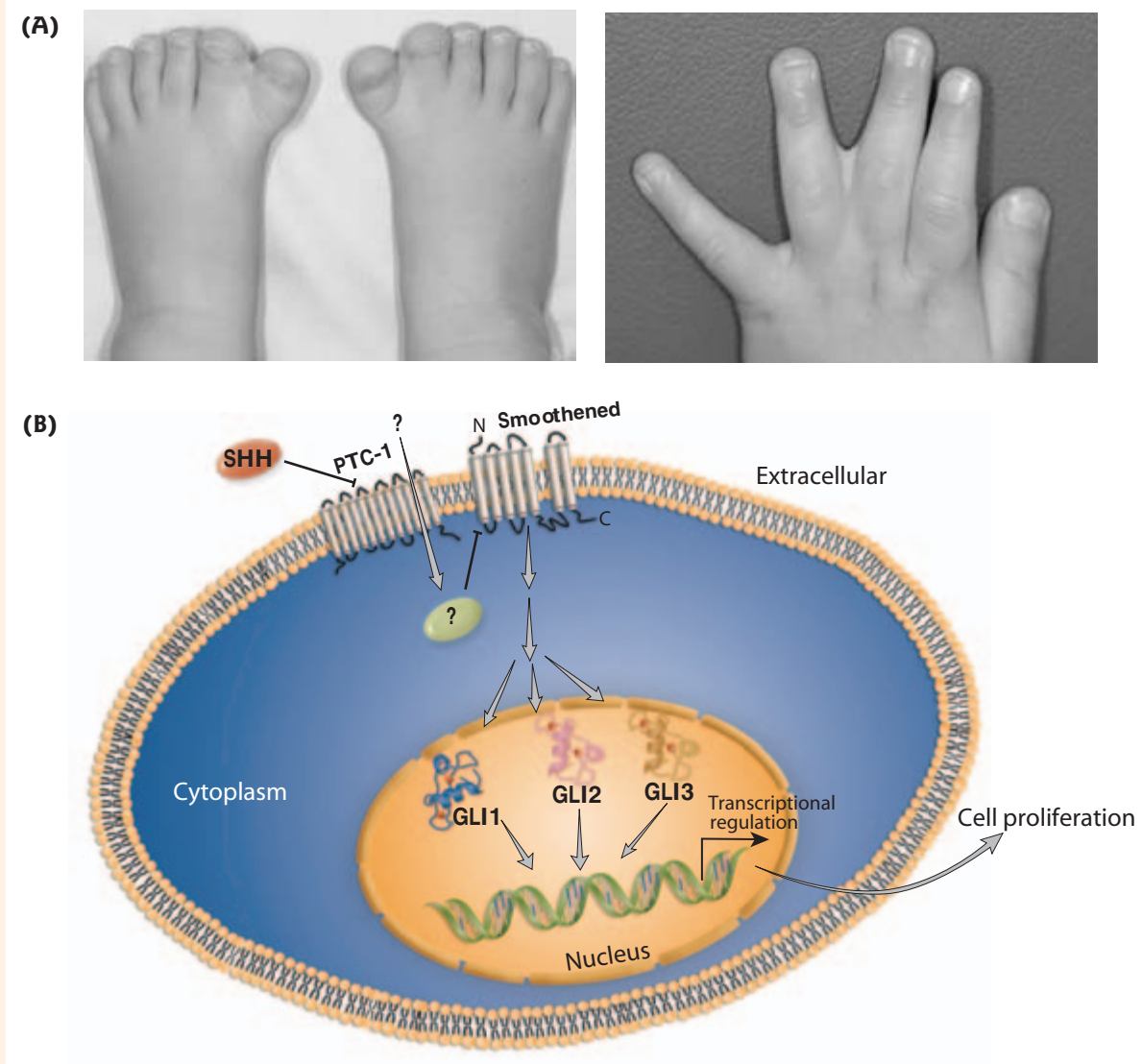
The Sonic hedgehog signal sets off a chain of events in target cells, leading to the activation and repression of target genes by transcription factors in the GLI family. In the presence of Sonic hedgehog, full-length GLI3 upregulates target genes, whereas in the absence of Sonic hedgehog, GLI3 is cleaved to produce a repressor that downregulates target genes. Sonic hedgehog is a secreted ligand that functions by binding to the transmembrane receptor proteins, Patched-1 and Patched-2. These receptors normally inhibit downstream signaling by interaction with Smoothened, another transmembrane receptor. When Sonic hedgehog is expressed, it relieves inhibition of Smoothened by Patched-1 and Patched-2, and allows Sonic hedgehog–Patched–GLI signaling to progress. Sonic hedgehog is an important signaling molecule for the developmental patterning of many tissues and organs. It is a key factor in the determination of the ventral neural tube, the anterior/posterior axis of the limb, and the cartilage-producing region of the somites. Studies of human GLI3 protein have shown that it interacts with CBP (CREB-binding protein), a common transcriptional coactivator. Mutations in CBP can cause Rubinstein–Taybi syndrome (Disease box 11.3). The similarity between the Rubinstein–Taybi phenotype and Greig cephalopolysyndactyly syndrome suggests that the biochemical interaction between CBP and GLI3 plays a role in Sonic hedgehog signaling.

The dimer can be either a homodimer – two of the same polypeptide – or a heterodimer – two different polypeptides zipped together. For example, C/EBP forms both homodimers and heterodimers composed of mixed pairs of different variants. The transcription factor AP-1 is a combination of members of two different families of transcription factors, Fos and Jun. Tony Kouzarides and Edward Ziff showed that at low concentrations of protein Jun and Fos bind to their DNA target better together than either one does separately (Fig. 11.18). They also demonstrated that the bZIP domains of each protein are essential for binding. Jun can form both homodimers and heterodimers, whereas Fos can only form heterodimers.

### Basic helix-loop-helix (BHLH)

The BHLH motif is distinct from the HTH motif (Fig. 11.17B). It forms two amphipathic helices, containing all the charged amino acids on one side of the helix, which are separated by a nonhelical





**Figure 1 Greig cephalopolysyndactyly syndrome and Sonic hedgehog signaling.** (A) Extra fingers and toes (polydactyly) and webbing and/or fusion of the fingers and toes (syndactyly) is a common malformation in Greig cephalopolysyndactyly syndrome due to a mutation in the *GLI3* gene which blocks Sonic hedgehog signaling. This patient's feet (left panel) show polysyndactyly, with left syndactyly of the second and third toes and right syndactyly of the first and second toes. The right hand panel shows syndactyly of the third and fourth fingers. (Reprinted with permission from Fujioka, H., Ariga, T., Horiuchi, K., Otsu, M., Igawa, H., Kawashima, K., Yamamoto, Y., Sugihara, T., Sakiyama, Y. 2005. Molecular analysis of non-syndromic preaxial polydactyly: preaxial polydactyly type-IV and preaxial polydactyly type-I. *Clinical Genetics* 67:429–433. Copyright © 2005 Blackwell Munksgaard.) (B) The Sonic hedgehog signal transduction pathway. The Sonic hedgehog signal sets off a chain of events in target cells, leading to the activation and repression of target genes by transcription factors in the GLI family, including GLI3. Sonic hedgehog (SHH) binds to the transmembrane receptor protein, Patched-1 (PTC-1). This receptor normally inhibits downstream signaling by interaction with Smoothened, another transmembrane receptor. When SHH is expressed, it relieves inhibition of Smoothened by PTC-1, and allows SHH–PTC-1–GLI signaling to progress. Not all the steps in the signaling pathway have been defined.



## DISEASE BOX 11.3

## Defective histone acetyltransferases in Rubinstein–Taybi syndrome

Rubinstein–Taybi syndrome is a rare disease that was first described by Rubinstein and Taybi in 1963. It is inherited in an autosomal dominant manner, and occurs approximately once in every 125,000 live births. The disease is characterized by facial abnormalities, broad digits, stunted growth, and mental retardation with impairments in learning and long-term memory formation. The disease is due to mutations in the gene coding for CBP which is a coactivator with histone acetyltransferase (HAT) activity. Patients have a variety of mutations in the CBP gene, including point mutations and 5′ or 3′ deletions. Patients are typically heterozygous for a mutation in CBP.

Homozygous mutations are likely to be embryonic lethal in humans.

The defect in long-term memory formation in patients is linked to the role of CBP in mediating transcription from some cAMP-responsive genes. The cAMP response element-binding protein (CREB) is a transcription factor that is activated by phosphorylation. Once activated, it recruits CBP and binds to a cAMP-responsive element (CRE) within the DNA. CBP acetylates the histones associated with the CRE, disrupting chromatin within the promoter and facilitating the binding of RNA pol II and the general transcription machinery.

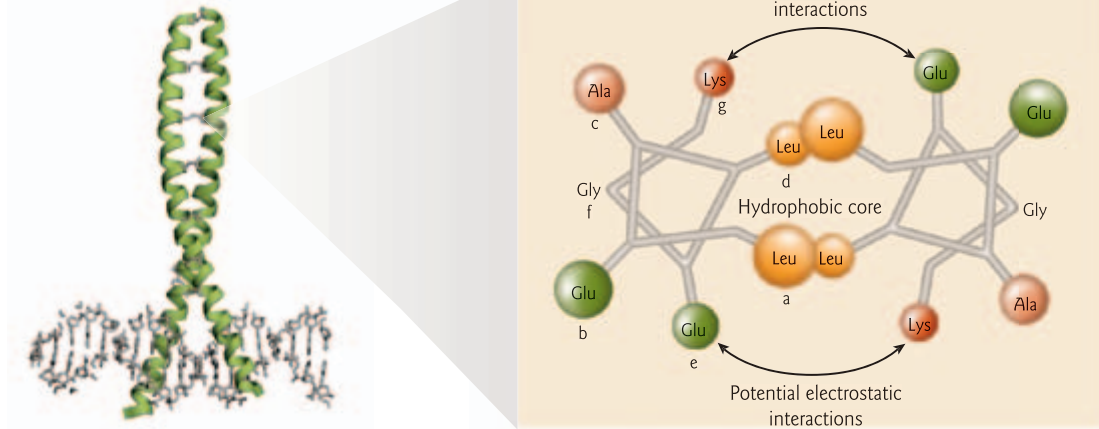
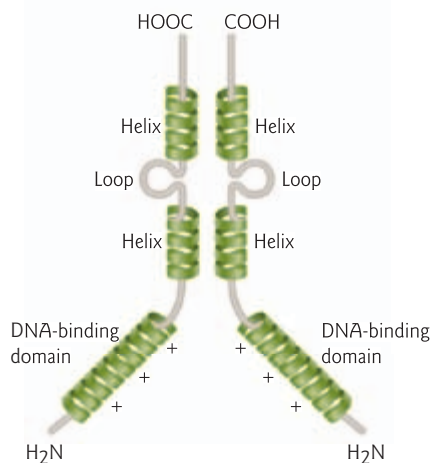
loop. Like the bZIP motif, the BHLH motif is not itself the DNA-binding domain of the transcription factor and does not directly contact the DNA. Instead, it plays an indirect structural role in DNA binding by facilitating dimerization of two similar transcription factors. The joining of two BHLH-containing transcription factors results in the correct positioning of the two adjacent DNA-binding domains in the dimer. The DNA-binding domains are rich in basic amino acids and can interact directly with the acidic DNA. As an example, efficient DNA binding by the mouse transcription factor Max requires dimerization with another BHLH protein. Max binds DNA as a heterodimer with either with Myc or Mad. The Myc–Max complex is a transcriptional activator, whereas the Mad–Max complex is a repressor.

### Transactivation domain

The transactivation domain of a transcription factor is involved in activating transcription via protein–protein interactions. Transactivation domains may work by recruiting or accelerating the assembly of the general transcription factors on the gene promoter, but their mode of action remains unclear. Some transcription factors do not contact the general transcription machinery directly but instead bind coactivators that in turn contact the general apparatus. Unlike the well-defined DNA-binding domains, transactivation domains are structurally more elusive. They are often characterized by motifs rich in acidic amino acids, so-called “acid blobs.” In addition to acid blobs, there are other distinct motifs. For example, transcription factor Sp1 contains a nonacidic transactivation region with multiple glutamine-rich motifs. Other motifs associated with transactivation include proline-rich regions and hydrophobic  $\beta$ -sheets.

### Dimerization domain

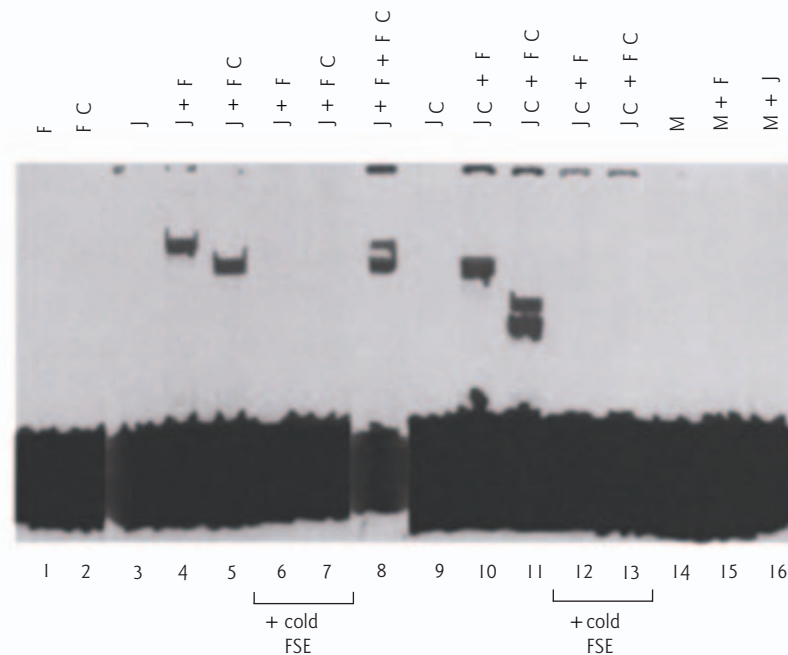
The majority of transcription factors bind DNA as homodimers or heterodimers. Accordingly, they have a domain that mediates dimerization between the two identical or similar proteins. In contrast to our detailed knowledge of protein–DNA interactions, far less is known about the exact molecular characteristics of these protein–protein contacts. As described above, two dimerization domains that are relatively well characterized structurally are the helix-loop-helix and leucine zipper motifs (Fig. 11.17).

**(A) Leucine Zipper (bZIP)****(B) Basic Helix-Loop-Helix (BHLH)**

**Figure 11.17 The basic DNA-binding domain.** (A) The leucine zipper and the basic DNA-binding domain are illustrated by the transcription factor AP-1–DNA complex. AP-1 is a dimer formed by Jun and Fos. Two helices “zipper” together by their leucine residues (inset) (Protein Data Bank, PDB: 1FOS.) (B) (Left) Schematic representation of the basic helix-loop-helix motif and the basic DNA-binding domain. (Right) Three-dimensional image of a Myc–Max–DNA complex. (Protein Data Bank, PDB: 1NKP).

## 11.6 Transcriptional coactivators and corepressors

Gene transcription is a multistep process involving a very large number of proteins functioning in discrete complexes. As described above, transcription factors bind to DNA in a sequence-specific manner. They mark a gene for activation or repression through the recruitment of coactivators or corepressors. Coactivators and corepressors are proteins that increase or decrease transcriptional activity, respectively, without binding DNA directly. Instead they bind directly to transcription factors and either serve as scaffolds for the recruitment of other proteins containing enzymatic activities, or they have enzymatic activities themselves for altering chromatin structure. Coactivators and corepressors have been much harder to study compared with transcription factors. In general, assays for protein–protein interactions are more difficult to perform than techniques for studying DNA–protein interactions (see Sections 9.8 and 9.9). In addition, techniques for determining which coactivator is docking on a particular transcription factor *in vivo* were not available until recently.



**Figure 11.18 Cooperative binding of Fos and Jun to DNA.** A  $^{32}\text{P}$ -labeled DNA that contains an AP-1 binding site was incubated with *in vitro*-translated products of Fos (F), Jun (J), the bZIP domains of Fos (Fos core, FC) and Jun (Jun core, JC), and Myc (M) in the combinations indicated at the top of each lane. Samples were then analyzed by an electrophoretic mobility shift assay (see Fig. 9.15A for methods). Lanes 6, 7, 12, and 13 contained excess unlabeled DNA competitor (+ cold FSE), which prevented interaction of the proteins with the labeled DNA. Combinations of Fos or FC with Jun or JC bound to the DNA (lanes 4, 5, 8, 10, and 11), but each of these proteins did not bind when added alone. The interaction was also specific because it did not occur with Myc, which is an unrelated bZIP protein. The conclusion was that Fos-Jun heterodimers bind better to AP-1-binding sites than Fos or Jun homodimers. (Reprinted by permission from Nature Publishing Group and Macmillan Publishers Ltd: Kouzarides, T., Ziff, E. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature* 336:646–651. Copyright © 1988.)

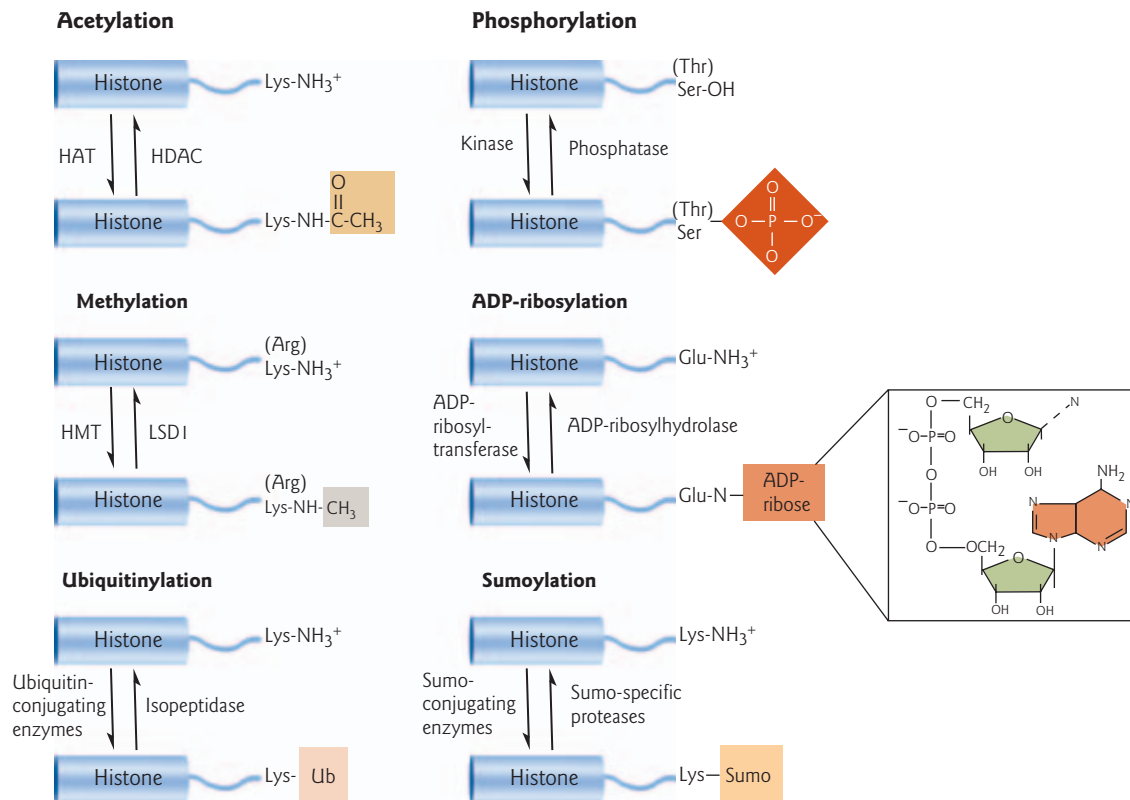
Coactivators, in the broadest sense, can be divided into two main classes:

- 1** Chromatin modification complexes: multiprotein complexes that modify histones post-translationally, in ways that allow greater access of other proteins to DNA.
- 2** Chromatin remodeling complexes: multiprotein complexes of the yeast SWI/SNF family (or their mammalian homologs BRG1 and BRM) and related families that contain ATP-dependent DNA unwinding activities.

Corepressors have the opposite effect on chromatin structure, making it inaccessible to the binding of transcription factors or resistant to their actions. Mechanisms for transcriptional silencing will be discussed in more detail in Chapter 12.

### Chromatin modification complexes

The assembly of the eukaryotic genome into chromatin is essential for the compaction of DNA into the relatively tiny nucleus (see Section 3.2). Until less than a decade ago, the histones were widely regarded as inert building blocks that package DNA into nucleosomes. Nucleosomes were viewed as general repressors

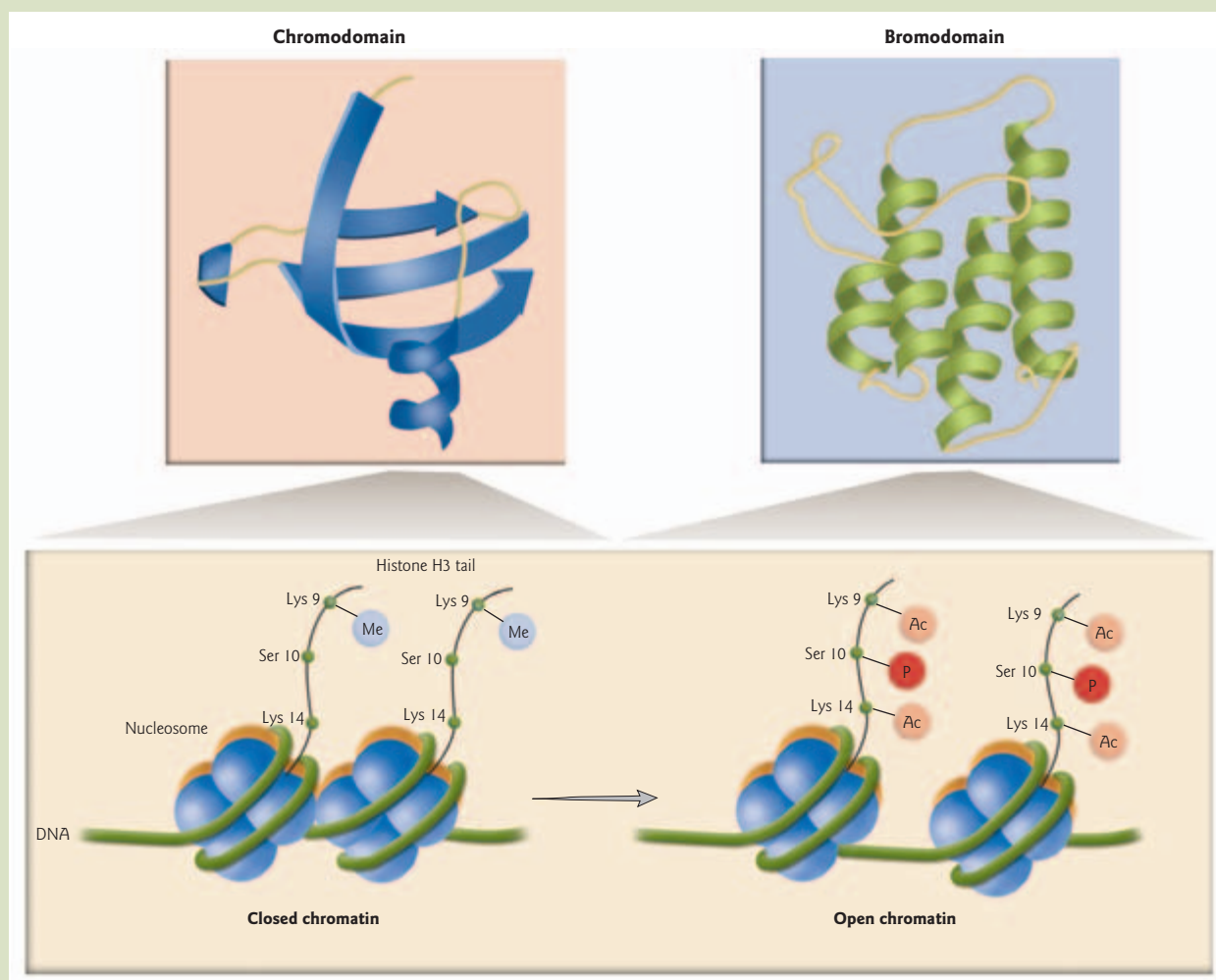


**Figure 11.19 Post-translational modification of histone N-terminal tails.** Six types of modifications to histone tails (wavy lines) known to play a role in gene regulation are depicted: acetylation, methylation, ubiquitinylation, phosphorylation, ADP-ribosylation, and sumoylation.

of transcription that make DNA sequences inaccessible to transcription factors. However, recent progress in the chromatin field has markedly changed our perspective. Chromatin modification complexes are now known to play a central role in both gene activation and repression.

The N-terminal tails of histones H2A, H2B, H3, and H4 stick out of the core octamer and are subject to a wide range of post-translational modifications (see Fig. 3.5). Because these covalent modifications alter the accessibility of chromatin to the general transcriptional machinery they have been proposed to function as master on/off switches that determine whether a gene is active or inactive. There are four main types of modification to histone tails known to play a role in regulating gene expression: (i) acetylation of lysines; (ii) methylation of lysines and arginine; (iii) ubiquitinylation of lysines; and (iv) phosphorylation of serines and threonines. Other less common modifications are ADP-ribosylation of glutamic acid and sumoylation of lysine residues. Levels of specific histone modifications or “marks” are maintained by the balanced activities of modifying and demodifying enzymes (Fig. 11.19). The activities of these two sets of enzymes may be shifted by changes in their intracellular distribution, their targeting to chromatin, or the action of inhibitors. In addition, some histone modifications may act as molecular switches, enabling or blocking the setting of other covalent marks. These modifications are used as recognition landmarks by other proteins that bind chromatin and initiate downstream processes, such as chromatin compaction or transcriptional regulation (Focus box 11.5). Besides modification, linker histone subtypes can also promote selective gene expression.





**Figure 1 Histone modifications are recognition landmarks for chromatin-binding proteins.** Proteins containing a chromodomain (left inset; Protein Data Bank, PDB: 1N72) are targeted to methylated lysines in histone tails, leading to tightly packed “closed” chromatin. Demethylation of the lysine amino acid at position 9 in histone H3 facilitates phosphorylation of serine 10 and the acetylation of lysines 9 and 15 leading to a more open chromatin structure. The bromodomain protein motif (right inset; PDB: 1PFB) forms a four-helix bundle with loops between pairs of helices. Conserved residues in these two loops form a hydrophobic pocket which serves as the binding site for an acetylated lysine.

### Histone acetyltransferases

The enzyme histone acetyltransferase (HAT) directs acetylation of histones. When a histone is acetylated, an acetyl group ( $\text{CH}_3\text{CO}$ ) is added to one or more of its lysine residues at the  $\epsilon$ -amino group, thereby removing the positive charge (Fig. 11.19). The addition of the negatively charged acetyl group reduces the

## Is there a histone code?

FOCUS BOX 11.5



The histone code hypothesis proposes that covalent post-translational modifications of histone tails are read by the cell and lead to a complex, combinatorial transcriptional output. In this model, the chromatin structure of a particular gene is specified by this combinatorial pattern of modifications of the histones that package it. Histone modifications are proposed to provide binding sites for proteins that can change the chromatin state to either active or repressed. Some protein domains present in chromatin-binding proteins and histone-modifying enzymes are known to interact with covalent marks in the histone tails. For example, the ~110 amino acid bromodomain recognizes specific acetylated lysines in histone tails. Another protein motif called the chromodomain (the name is derived from “chromatin organization modifier”) is targeted to methylated lysines and arginines in histone tails (Fig. 1).

An increasing amount of experimental data has provided support for different aspects of the histone code hypothesis, but the hypothesis continues to be the subject of much debate. To date, acetylation of individual lysines has not been shown to be combinatorial or consistent from gene to gene. Genome-wide analysis of histone acetylation patterns shows evidence for groups of genes with consistent acetylation patterns involving combinations of 11 lysines. However, some researchers argue that the use of histone modifications

individually or sequentially cannot be considered a code since the total number of modifications does not necessarily contain more information than the sum of the individual modifications. Only if acetylation patterns are read combinatorially can they provide the language of a true code. Single or sequential use of modifications is proposed to be more analogous to a protein-signaling pathway.

Recently, a detailed study of the complexity of histone H4 acetylation patterns was performed that provides some insight into the nature of the “code.” Researchers constructed yeast strains carrying all possible combinations of mutations among four lysines in the histone H4 tail that are involved in transcriptional silencing when acetylated. They then characterized the resulting genome-wide changes in gene expression by using DNA microarrays (see Fig. 16.13). They found that the four H4 lysines give rise to eight transcriptional states, rather than the full complement of 16 possible states. Cumulative effects were seen as changes in gene expression correlating with an increase in the total number of mutations; the consequences of H4 acetylation are simple and cumulative. The researchers concluded that, at least for these particular histone modifications, if there is indeed a “code” it is a simple one and not a “complex, multimark code.”

overall positive charge of the histones. This results in a decreased affinity of the histone tails for the negatively charged DNA. Histone deacetylase (HDAC) catalyzes the removal of acetyl groups from lysines near the N-termini of histones. Because of the decreased affinity of acetylated histone tails for DNA, hyperacetylation of histones was generally assumed to be associated with gene activation. This view has been modified by two lines of evidence. First, genome-wide analysis indicates that both acetylation and deacetylation of specific lysines are correlated with gene activity. Second, histone deacetylation by a particular HDAC (Hos2) was shown to be required for the activity of specific genes. The current model proposes that the acetylation state of lysine residues provides a specific binding surface that can either serve to recruit repressors or activators of gene activity, depending on the context.

### Histone methyltransferases

The enzyme histone methyltransferase (HMT) directs methylation of histones. Unlike histone acetylation which takes place only on lysine (K), methylation occurs both on lysine and arginine (R) residues. Either one, two, or three methyl groups ( $\text{CH}_3$ ) are added at lysine residues. The methyl groups increase their bulk but do not alter the electric charge (Fig. 11.19). Arginine can be modified by the addition of one methyl

group. Histone methylation is linked to both activation and repression. The site specificity of lysine or arginine methylation, as well as the number of methyl groups attached to a particular lysine or arginine, can have distinct effects on transcription. For example, dimethylation of histone H3 at lysine 9 (abbreviated as dimethylated H3-K9) and trimethylation of histone H3 at lysine 27 (trimethylated H3-K27) are both largely associated with gene silencing and heterochromatin formation. In contrast, methylation of H3-K4, H3-K36, or H3-K79 is associated with active chromatin. Demethylation of H3-K9 facilitates phosphorylation at serine 10 and acetylation at lysine 14. In addition, serine 10 phosphorylation facilitates acetylation of lysine 9, thereby preventing the setting of the repressive lysine 9 methylation mark (see Focus box 11.5). These modifications lead to tightly packed, “closed” chromatin becoming “open” chromatin. Histone methylation is generally associated with the recruitment of other proteins, which may then combine to form an altered chromatin structure.

Histone acetylation was known to be dynamically regulated by HATs and HDACs, but histone methylation had been considered to be a “permanent” modification, because no histone demethylase had been isolated. However, in 2004, scientists finally isolated an enzyme that acts as a histone demethylase. The enzyme – lysine-specific demethylase 1 (LSD1) – represses specific genes by maintaining unmethylated histones. LSD1 is conserved from *Schizosaccharomyces pombe* to humans. Discovery of this enzyme suggests that histone methylation is also a dynamic process and is subject to regulation by both methylases and demethylases (see Fig. 11.19).

### Ubiquitin-conjugating enzymes

Usually, the addition of polyubiquitin chains targets a protein for degradation by the proteasome (see Fig. 5.17). However, the addition of one ubiquitin (monoubiquitinylation) can alter the function of a protein without signaling its destruction. A conjugating enzyme catalyzes the formation of a peptide bond between ubiquitin and the side chain-NH<sub>2</sub> of a lysine residue in a target protein. For example, monoubiquitinylation of histone H2B is, depending on the gene, associated with activation or silencing of transcription as well as transcription elongation. Monoubiquitinylation of linker histone H1 leads to its release from the DNA. In the absence of the linker histone, chromatin becomes less condensed, leading to gene activation.

### Kinases

When a histone is phosphorylated by a specific kinase, a phosphate group is added, usually to one or more of its serine or threonine amino acids, thereby adding a negative charge (see Fig. 11.19). Phosphorylation of the linker histone H1 either removes H1 from the DNA, or makes it bind DNA less tightly. Either case is generally associated with gene activation. As noted above, phosphorylation of histone H3 is also associated with the activation of specific genes. Phosphorylation is a dynamic process, subject to regulation by phosphatases.

### ADP-ribosyltransferases

Mono-ADP-ribosylation is the enzymatic transfer of an ADP-ribose residue from NAD<sup>+</sup> to a specific amino acid of the acceptor protein by an ADP-ribosyltransferase (see Fig. 11.19). Histones are modified at glutamic acid residues; however, in other proteins cysteine, asparagine, and arginine residues can be modified. Removal of the ADP-ribose residue is mediated by an ADP-ribosylhydrolase. MacroH2A, a histone variant associated with X chromosome inactivation (see Section 12.4), has an N-terminal region with high sequence homology to H2A, but it also contains a 25 kDa nonhistone macro domain. A recent crystallographic study points to the possibility that macro domains may function in ADP-ribosylation of histones. This was an exciting finding because it suggests for the first time that a histone variant may have inherent enzymatic potential.

### SUMO-conjugating enzymes

Small ubiquitin-like modifier (SUMO) has 20% identity with ubiquitin. SUMO conjugation involves the same set of enzymatic steps as ubiquitin conjugation (see Fig. 5.17). SUMO is a 97 amino acid protein that is added through its carboxyl end to the amino functional group on lysines in a target protein (see Fig. 11.19). Typically, lysine residues subject to sumoylation are found within a consensus motif,  $\Psi$ -Lys- $x$ -Glu (where  $\Psi$  is a large hydrophobic residue and  $x$  is any residue). Deconjugation is mediated by SUMO-specific proteases. In most cases reported to date, sumoylation is associated with transcriptional repression. Once thought to be a relatively rare post-translational modification, sumoylation is now known to play a role in regulating protein transport, segregation of chromosomes during mitosis, and repair of damaged DNA. Novel functions of SUMO continue to be discovered. Although SUMO can operate throughout the cell, its main actions are concentrated in the nucleus.

### Linker histone variants

In some cases, the selection of a different histone linker protein is important for gene regulation. Mammals contain eight histone H1 subtypes including H1a through H1e and H1<sup>o</sup> found in somatic cells, as well as two germ cell-specific subtypes, H1<sub>t</sub> and H1<sub>oo</sub>. The expression of each mammalian histone H1 subtype depends on the tissue type, phase of the cell cycle, and developmental stage. For example, when the Msx1 homeodomain protein forms a complex with linker histone H1b, the complex binds to an enhancer element in the *MyoD* gene and inhibits gene expression. This prevents the differentiation of muscle progenitor cells because MyoD (a helix-loop-helix transcription factor) is a master regulator of skeletal muscle cell differentiation. Decreased production of histone H1b and Msx1 during development relieves the repression of the *MyoD* gene, allowing skeletal muscle differentiation to proceed (Fig. 11.20).

### Chromatin remodeling complexes

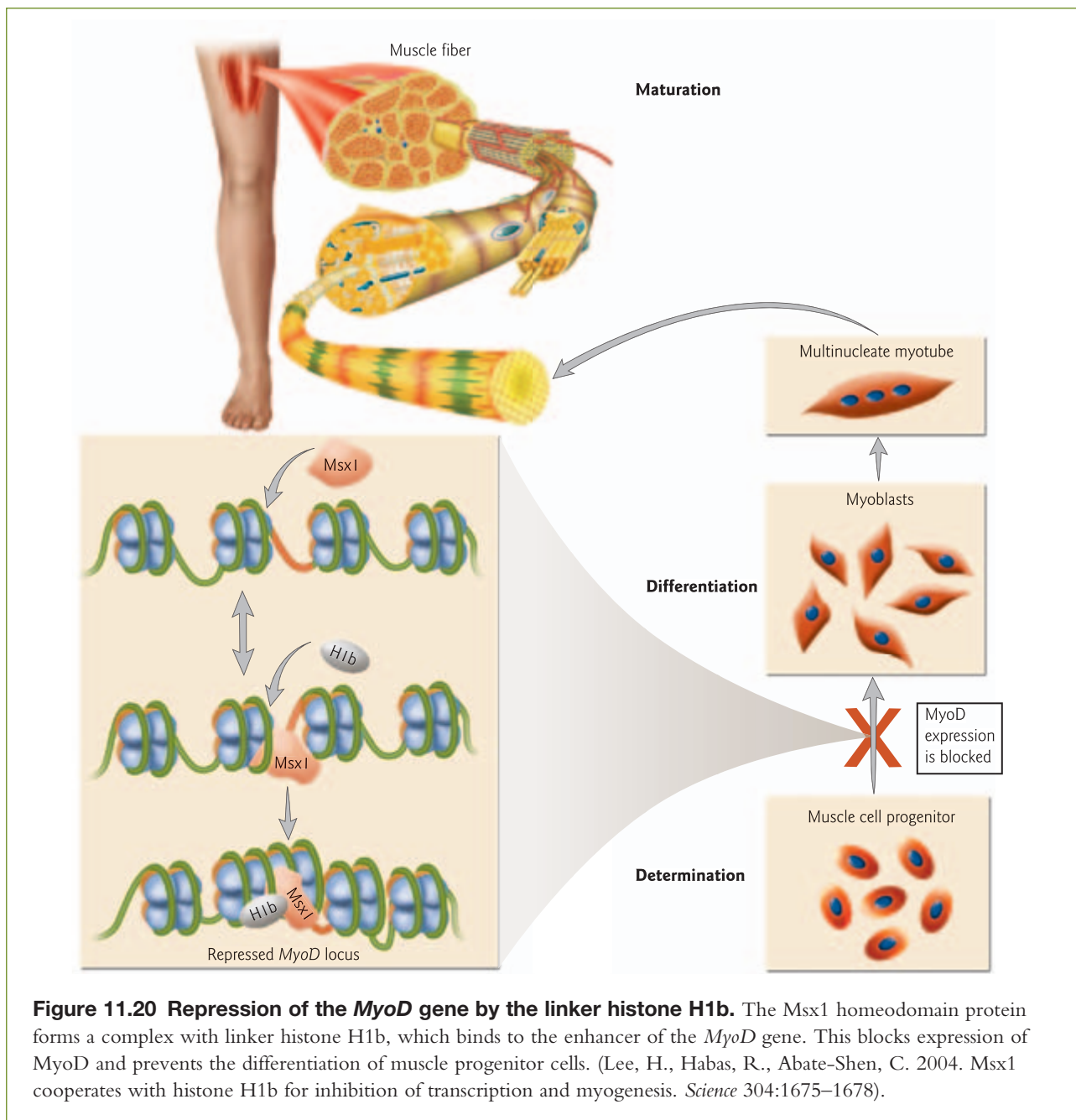
The recent discovery of ATP-dependent chromatin remodeling complexes marked an exciting advance in the chromatin field. This new class of molecular motors uses the energy derived from ATP hydrolysis to change the contacts between histones and DNA, thereby allowing transcription factors to bind to DNA regulatory elements. Chromatin remodeling complexes can mediate at least four different changes in chromatin structure (Fig. 11.21):

- 1 Nucleosome sliding: the position of a nucleosome changes on the DNA.
- 2 Remodeled nucleosomes: the DNA becomes more accessible but the histones remain bound.
- 3 Nucleosome displacement: the complete dissociation of DNA and histones.
- 4 Nucleosome replacement: replacement of a core histone with a variant histone.

There are a number of different families of ATP-dependent remodeling complexes that have been characterized, including the SWI/SNF, ISWI, and SWR1 complex families. Each family is defined by a unique subunit composition and the presence of a distinct ATPase. The precise mechanism of remodeling appears to be distinct from one family member to another.

### SWI/SNF chromatin remodeling complex family

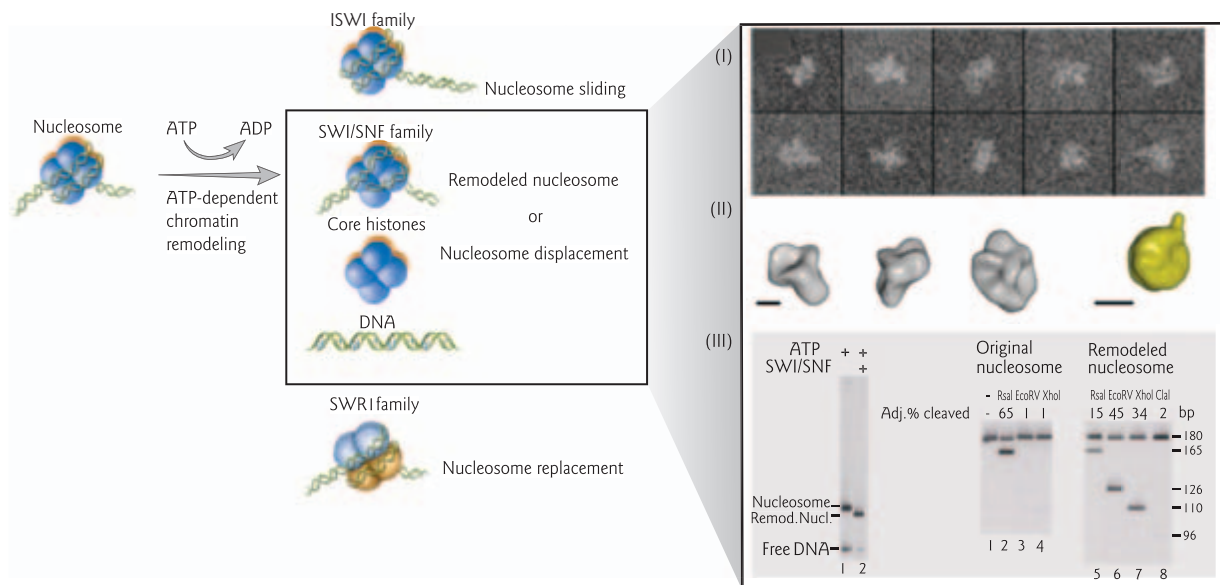
The SWI/SNF complex (pronounced “switch-sniff”) from the budding yeast *Saccharomyces cerevisiae* was the first chromatin remodeling complex to be characterized. It is a massive 2 MDa complex composed of at least 11 different polypeptides (Fig. 11.21). The name derives from the components of the catalytic ATPase subunit, which were first described in mutants that were mating-type *switching* defective (Swi2) (see Section 12.7) and sucrose *nonfermenters* (Snf2). Many other chromatin remodeling factors have since been identified. The human cell homolog of SWI/SNF is called BAF (*BRG-1 associated factor*) and



contains two ATPase subunits, human Brahma (hBRM) and Brahma-related gene 1 (BRG1). The *Drosophila* homolog BAP (BRM complex associated protein) has a single ATPase subunit named Brahma.

RSC (remodel the structure of chromatin) is another yeast chromatin remodeling complex closely related to SWI/SNF. This complex contains about 15 subunits, sharing two identical subunits and at least four homologs with the subunits of SWI/SNF. PBAF and PBAP are RSC-like complexes found in mammals and *Drosophila*, respectively. The “P” stands for “polybromo associated.” In polybromo-associated proteins there is an unusually high number of bromodomains (Focus box 11.5). Typically bromodomains are only present as single or double domains, but in PBAP there are six bromodomains. The bromodomain is a distinguishing feature of the SWI/SNF-like family that is absent in ISWI and other remodeling complexes.

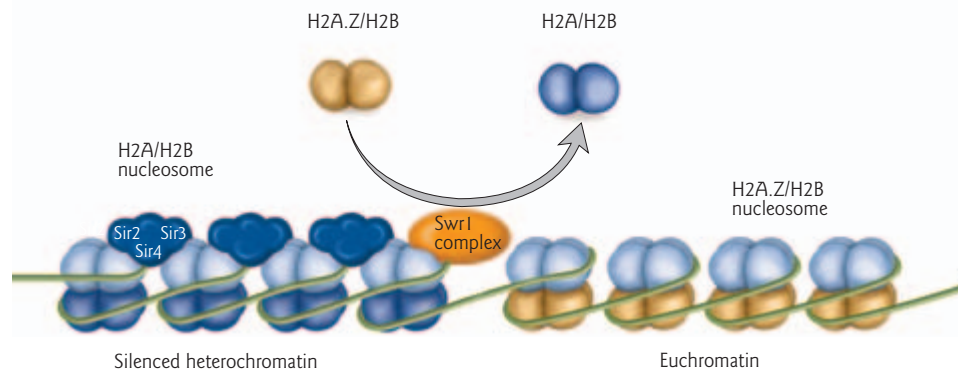




**Figure 11.21 Various modes of ATP-dependent chromatin remodeling.** ATP-dependent chromatin remodeling complexes alter histone–DNA contacts. Nucleosomes either slide to another position (ISWI family), or are remodeled or displaced (SWI/SNF family), or core histones are replaced with a variant (SWR1 family). (Inset) (i) Raw images of the yeast SWI/SNF complex obtained by scanning transmission electron microscopy. (ii) Rendered surface three-dimensional structures of the SWI/SNF complex (gray) and the nucleosome core particle (yellow) for comparison. The SWI/SNF complex has a cone-shaped depression at the top which may serve as a nucleosome-binding pocket. (Reprinted with permission from: Smith, C.L., Horowitz-Scherer, R., Flanagan, J.F., Woodcock, C.L., Peterson, C.L. 2003. Structural analysis of the yeast SWI/SNF chromatin remodeling complex. *Nature Structural Biology* 10:141–145. Copyright © 2003 Nature Publishing Group). (iii) Remodeled nucleosomes have faster mobility in an electrophoretic mobility shift assay (EMSA) (see Fig. 9.15A) and differential accessibility to restriction endonuclease digest (see Section 8.3). Core octamers were assembled *in vitro* on DNA radiolabeled at the 5' end, and incubated with or without SWI/SNF. After remodeling, SWI/SNF was removed from the labeled nucleosomes by the addition of excess unlabeled competitor DNA. Samples were separated by native polyacrylamide gel electrophoresis (PAGE). A small proportion of nucleosomes were released as a slightly slower migrating band, identified as a hexamer–DNA complex. Nucleosomes were purified from the gel and digested with various restriction endonucleases as indicated. The resulting restriction fragments were separated by denaturing PAGE. (Reprinted from: Kassabov, S.R., Zhang, B., Persinger, J., Bartholomew, B. 2003. SWI/SNF unwraps, slides, and rewraps the nucleosome. *Molecular Cell* 11:391–403. Copyright © 2003, with permission from Elsevier.)

### Mode of action of SWI/SNF: nucleosome sliding and disassembly

The SWI/SNF and RSC-type chromatin remodeling complexes cause nucleosome sliding and significant perturbation of nucleosome structure. A general model of the mode of action of SWI/SNF is shown in Fig. 11.21. The ATPase activity of SWI/SNF hydrolyzes ATP and the energy is used for modification of the path of DNA around the histone octamer. Some 50 bp of DNA is unwrapped from the edges of the nucleosome. This results in the exposure of DNA on the nucleosomal surface and the sliding of nucleosomes to new positions. In fact, in some highly active genes, such as the yeast *PHO5* gene, the nucleosomes dissociate completely from the DNA. *PHO5* encodes a secreted acid phosphatase and is induced in response to a lack of inorganic phosphate in the growth medium. After the remodeling complex alters the path of DNA around the histone octamer, transcription factors and the general transcription machinery can gain access to DNA. Often there is cooperation between SWI/SNF and histone acetyltransferase (HAT) in remodeling chromatin structure.



**Figure 11.22 The SWR1 complex mediates replacement of a core histone with a variant histone.**

H2A-H2B dimers (light blue) are replaced with H2A.Z-H2B dimers (brown) by the action of the yeast SWR1 complex. The variant histone H2A.Z prevents the spread of Sir proteins (dark blue) associated with silenced chromatin (heterochromatin) at telomeres into adjacent regions of euchromatin. (Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., Wu, C. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303:343–348).

### ISWI chromatin remodeling complex family

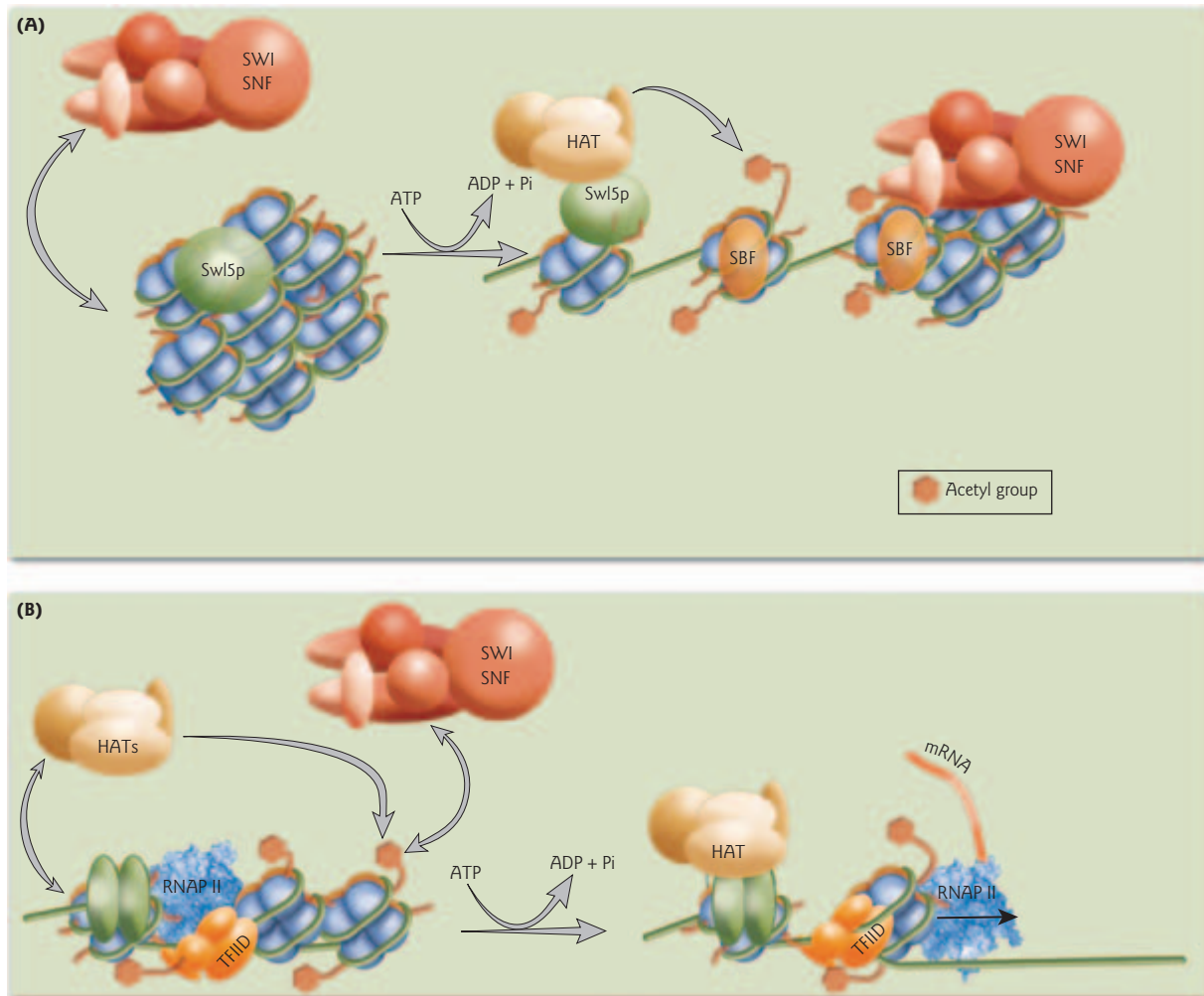
In contrast to SWI/SNF and RSC-type chromatin remodeling complexes, members of the ISWI (*imitation Swi2*) family (such as NURF from *Drosophila*) relocate the nucleosomes by sliding the histone octamers along the DNA without apparent perturbation of their structure (see Fig. 11.21). This process yields nucleosomes at different positions on the DNA without their disassembly, or the altered nucleosome forms that are typical of SWI/SNF action. Evidence for this comes from experiments demonstrating that the ISWI complexes do not alter the nuclease sensitivity of nucleosome core particles.

### SWR1 chromatin remodeling complex family

The SWR1 complex is named after its ATPase subunit Swr1 (for *Swi2/Snf2* related). SWR1 may be targeted to the appropriate genomic regions by the Bdf1 subunit (*bromodomain factor 1*), which contains two bromodomains. The SWR1 complex adds a new theme to chromatin remodeling: histone replacement with a variant histone in the core octamer (Fig. 11.22). The disruption of the core octamer was demonstrated by studies showing that the SWR1 complex can transfer H2A.Z-H2B dimers in exchange for H2A-H2B dimers to nucleosomal arrays but not to free DNA. The reaction is dependent on the ATPase subunit of SWR1. Histone variant H2A.Z is strongly enriched in active genes that are adjacent to transcriptionally repressed regions, such as near telomeres. The presence of this variant histone prevents the spread of silent heterochromatin into neighboring euchromatin.

## 11.7 Transcription complex assembly: the enhanceosome model versus the “hit and run” model

Knowledge of the identity and roles of transcription factors and coactivators is well advanced. However, the dynamic process by which these proteins interact on DNA to activate transcription is the subject of much study. There is clearly an interplay between chromatin remodeling and modification complexes. Both types of coactivators can be recruited to the same gene promoters, but there appears to be no general rule for the order of recruitment of the many different proteins involved in gene transcription. Instead there is a gene-specific order of events.



**Figure 11.23 The order of recruitment of transcriptional regulatory proteins is gene-specific.** (A) At the yeast *HO* gene promoter, the Swi5p transcription factor recruits SWI/SNF and a histone acetyltransferase (HAT) complex, followed by a second transcription factor, SBF, well before assembly of the preinitiation complex. (B) At the human  $\alpha$ -antitrypsin gene promoter, multiple HAT complexes and the SWI/SNF complex are recruited after preinitiation complex assembly. (Reprinted with permission from: Fry, C.J., Peterson, C.L. 2002. Unlocking the gates to gene expression. *Science* 295:1847–1848. Copyright © 2002 AAAS.)



## Order of recruitment of various proteins that regulate transcription

The order of recruitment depends on chromatin structure of the gene promoter, the phase of the cell cycle, and many other factors. Two examples are shown in Fig. 11.23. In the first example, SWI/SNF is recruited by an activator protein bound to an upstream regulatory region in the yeast *HO* gene (see Section 12.7). After SWI/SNF remodels the *HO* gene promoter, a histone acetyltransferase (HAT) complex is recruited. The cooperative action of SWI/SNF and HAT facilitates the binding of a gene-specific transcription factor, followed by the general transcription factors and RNA pol II. Unexpectedly, in the second example, the entire preinitiation complex is assembled on the  $\alpha_1$ -antitrypsin gene promoter prior to recruitment of chromatin remodeling complexes or HATs. The human interferon- $\beta$  (IFN- $\beta$ ) gene promoter follows a different series of events upon induction by viral infection. A DNA-activator complex (enhanceosome, see

below) binds to a nucleosome-free region upstream of the IFN- $\beta$  gene promoter. This leads to rapid recruitment of HAT and acetylation of nucleosomes at the TATA box in the gene promoter. Subsequently, the SWI/SNF complex slides a nucleosome array from the TATA box, allowing access to the general transcription machinery.

Individual transcription factors generally bind to DNA regulatory elements such as promoters and enhancers with relatively low specificity both *in vitro* and *in vivo*. However, a high degree of specificity is achieved through combinatorial interactions that occur when multiple transcription factors bind to clustered DNA regulatory elements. Cooperative binding of regulatory proteins to DNA was introduced as an important mechanism for bacterial gene regulation in Chapter 10 (Section 10.6), and is an equally important mechanism in eukaryotic gene regulation. Two models for binding of transcription factors and assembly of transcription complexes have been proposed, the enhanceosome model and the “hit and run” model. A recent study suggests that these models are not mutually exclusive.

### Enhanceosome model

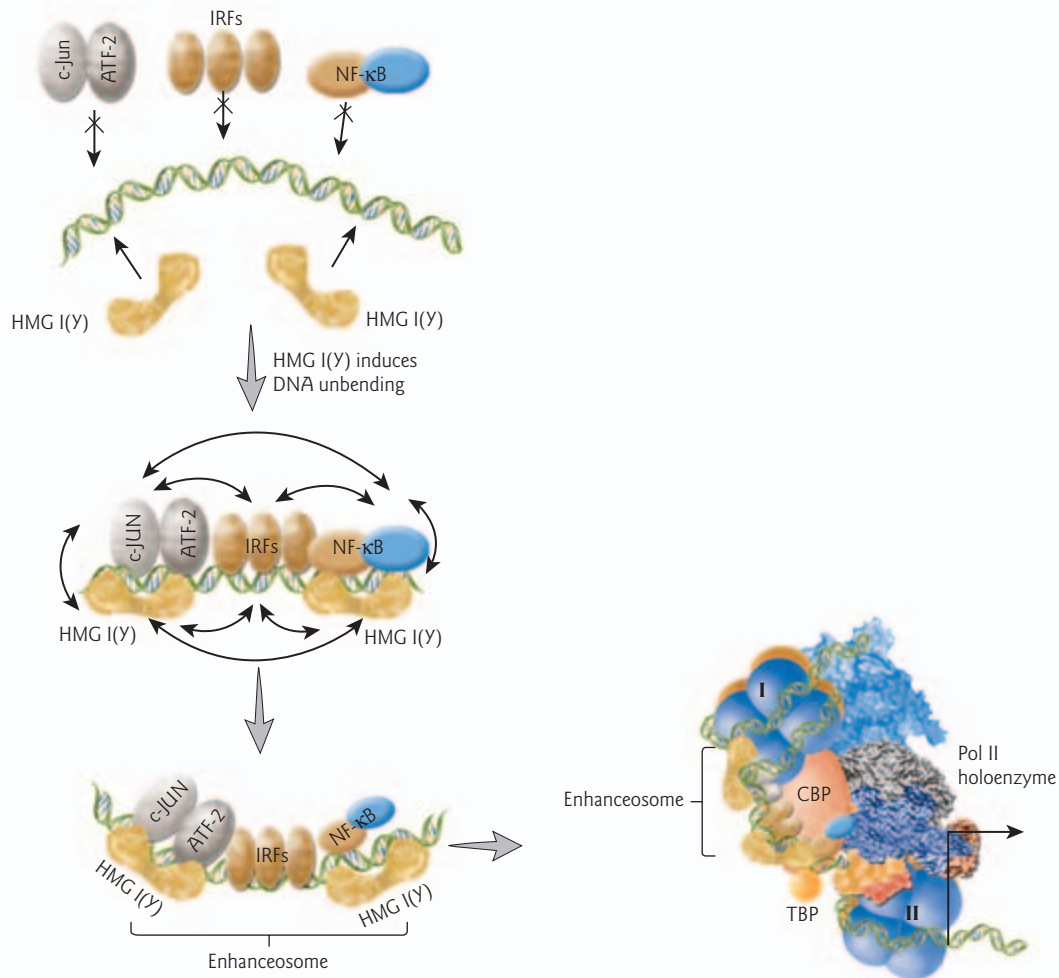
The enhanceosome model proposes that interactions among transcription factors promote their cooperative, step-wise assembly on DNA and give the complex exceptional stability. The context in which the elements are organized within three-dimensional space is essential for transcriptional regulation, as are extensive protein–protein and DNA–protein interactions. This model is compatible with transcription complex assembly within the IFN- $\beta$  gene enhancer. The inducible enhancer consists of multiple binding sites for transcription factors NF- $\kappa$ B (see Section 11.10), IRF1 (*interferon regulatory factor 1*), activating transcription factor 2 (ATF2), and c-Jun. Although each site will not activate transcription independently, together they facilitate the activation process. For example, HMG-I/Y (*high mobility group protein isoform I and Y*) is essential for directing the appropriate unbending of the IFN- $\beta$  gene enhancer to form the “enhanceosome” (Fig. 11.24). Members of the HMG family are “architectural proteins” that are the major nonhistone component of mammalian chromatin. They bind AT-rich DNA within the minor groove and induce conformational changes in the DNA double helix that facilitate the formation of multiprotein complexes. By binding to DNA near the NF- $\kappa$ B- and ATF2/c-Jun-binding sites, HMG-I/Y makes direct protein–protein contacts with the two transcription factors and increases their affinity for DNA. Helical-twist experiments (e.g. Fig. 10.18) show that the relative phasing of transcription factor and HMG-I/Y-binding sites cannot be changed without interfering with the induction of transcription.

### Hit and run model

The classic enhanceosome model at first glance seems incompatible with the observation that transient and dynamic binding is a common property of all chromatin proteins with the exception of core histones. Fluorescence recovery after photobleaching (FRAP) experiments (Fig. 11.25A) have shown that most nuclear proteins are highly mobile and the interaction of proteins with chromatin and nuclear compartments is highly dynamic. For example, the glucocorticoid receptor was shown by FRAP to bind and unbind to chromatin in cycles of only a few seconds. Even such transient interactions, however, are sufficient to promote large-scale remodeling of chromatin lasting a few hours and to correlate with transcriptional activation. In the hit and run model, transcriptional activation reflects the probability that all components required for activation will meet at a certain chromatin site (the “hit”) – i.e. transcription complexes are assembled in a stochastic fashion from freely diffusible proteins – and that their binding is transient (the “run”).

### Merging of models

A more recent study suggests that the two models are not mutually exclusive. Each protein in a complex influences the binding kinetics of its partners. The bottom line is that the principles of combinatorial interaction and complex stability apply to hit and run models even if the complex itself has a very limited

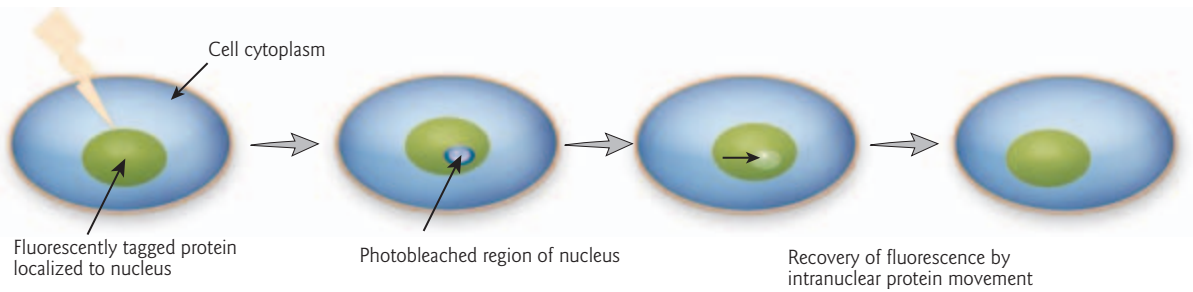


**Figure 11.24 The INF- $\beta$  enhanceosome.** A two-step model for enhanceosome assembly. Transcription factors recognize their binding sites in the INF- $\beta$  promoter region with low affinity due to unfavorable intrinsic DNA curvature. Binding of HMG-I/Y to the promoter region unbends DNA, lowering the free energy required for binding of the transcription factors c-Jun, ATF2, IRF1, and NF- $\kappa$ B. Enhanceosome assembly is completed by protein–protein interactions (arrows) between all the components leading to a highly stable structure. The enhanceosome interacts with the general transcription machinery at the core promoter through protein–protein interactions. Multiple HATs are recruited during assembly of the preinitiation complex, which includes RNA polymerase II (Pol II holoenzyme). The SWI/SNF complex joins the assemblage and facilitates TBP binding to the TATA box. Nucleosomes I and II have been remodeled by the SWI/SNF complex. CBP is a histone acetyltransferase complex. The black arrow indicates the start of transcription. (Yie, J., Merika, M., Munshi, N., Chen, G., Thanos, D. 1999. The role of HMG I(Y) in the assembly and function of the INF- $\beta$  enhanceosome. *EMBO Journal* 18:3074–3089; and Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., Thanos, D. 2000. Ordered recruitment of chromatin modifying and general transcription factors to the IFN- $\beta$  promoter. *Cell* 103:667–678.)

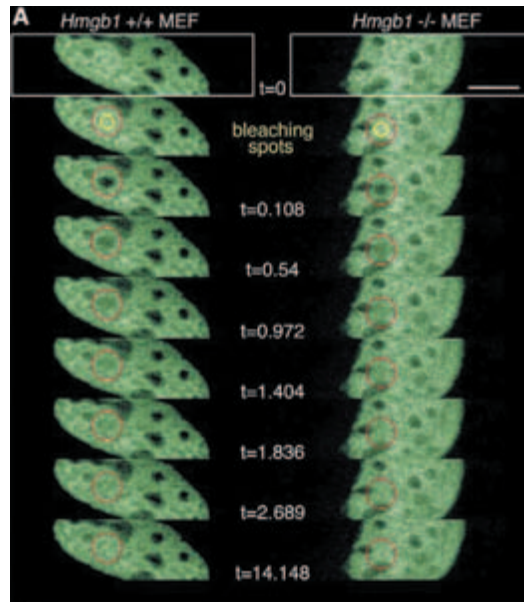
lifetime. High mobility group box 1 protein (HMGB1) and the glucocorticoid receptor (GR) were analyzed in living cells, by tagging them with yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP), respectively (see Section 9.3). The interaction between CFP–GR and HMGB1–YFP in living cells was followed via fluorescence resonance energy transfer (FRET) (see Fig. 9.16D) and FRAP. The two proteins were shown to only interact within chromatin and to lengthen each other's residence time on chromatin



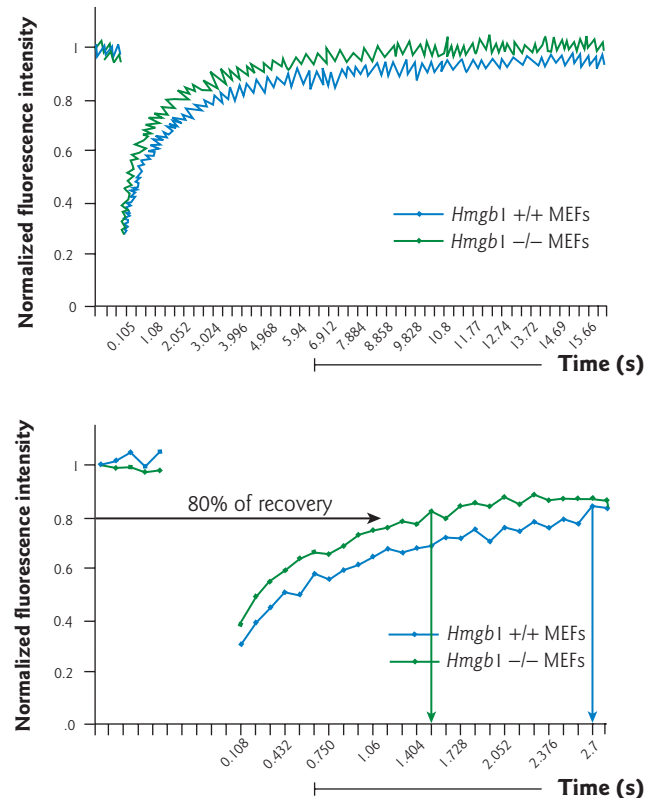
**(A) Laser bleach pulse**



**(B)**



**(C)**

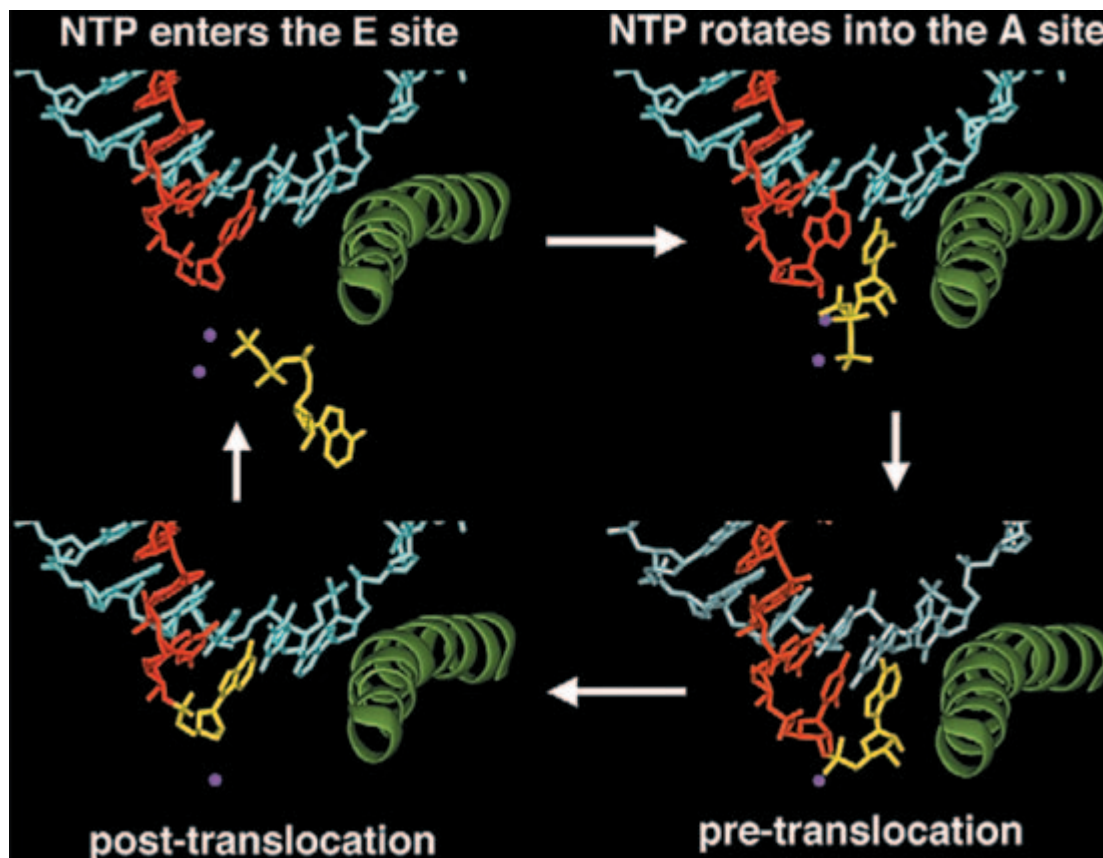


**Figure 11.25 Protein dynamics within the nucleus.** (A) In a fluorescence recovery after photobleaching (FRAP) experiment, a small area of the nucleus in a cell expressing a fluorescently tagged protein is rapidly and irreversibly bleached using a targeted laser pulse. Photobleaching results in a small region devoid of fluorescence signal. The recovery of the fluorescence signal is measured as a function of time using confocal microscopy. Recovery of fluorescence is due to the movement of unbleached proteins into the bleached area. The kinetics of recovery correlates with the mobility of the fluorescently tagged proteins. (B) GFP-tagged glucocorticoid receptor (GR) shows increased mobility in *Hmgbl*<sup>-/-</sup> knockout mouse cells (see Fig. 15.4 for methods) as compared to *Hmgbl*<sup>+/+</sup> wild-type mouse cells. Yellow circles indicate the photobleaching areas in the nucleus (green), and the red contours highlight the region of interest. Images were collected every 108 ms. The scale bar represents 10  $\mu$ m. (C) FRAP analysis of GFP-GR in *Hmgbl*<sup>-/-</sup> (green dots) and *Hmgbl*<sup>+/+</sup> (blue dots) mouse cells. (Upper graph) The dots indicate mean values, and error bars indicate standard error, from 10 cells of a representative experiment. (Lower graph) An enlargement of the recovery kinetics of the two cells closest to the mean in their group, in order to show the calculation of 80% recovery. Recovery is faster in knockout cells than in wild-type cells: 80% recovery is reached in 2.7 and 1.4 seconds respectively; complete recovery in knockout cells is reached in 9 seconds; while it is beyond 16 seconds in wild-type cells. (Reprinted from: Agresti, A., Scaffidi, P., Riva, A., Caiola, V.R., Bianchi, M.E. 2005. GR and HMGB1 interact only within chromatin and influence each other's residence time. *Molecular Cell* 18:109–121. Copyright © 2005, with permission from Elsevier.)

(Fig. 11.25B). Mutants that were unable to bind chromatin showed no FRET, and GR was more mobile in cells devoid of HMGB1 (e.g. *Hmbg1*<sup>-/-</sup> cells from knockout mice).

## 11.8 Mechanism of RNA polymerase II transcription

Crystallographic models at 2.3 Å of the complete RNA pol II, together with new biochemical and electron microscopic data, have begun to provide mechanistic insights into the process of initiation and subsequent synthesis of the RNA transcript. After assembly of the preinitiation complex, a period of abortive initiation follows before the polymerase escapes the promoter region (promoter clearance) and enters the elongation phase. The basic process of RNA synthesis is subdivided into multiple stages: selection of a nucleoside triphosphate (NTP) complementary to the DNA template, catalysis of phosphodiester bond formation, and translocation of the RNA and DNA (Fig. 11.26). In essence, the overall process is the same as that described



**Figure 11.26 RNA transcript synthesis by RNA pol II.** Four crystal structures of RNA pol II transcribing complexes are shown illustrating the four step cycle of RNA transcript synthesis. (1) A nucleotide (NTP) enters the entry (E) site beneath the active center. (2) The NTP rotates into the nucleotide addition (A) site and is checked for mismatches. (3) Pretranslocation: phosphodiester bond formation. (4) Translocation and post-translocation: the NTP just added to the RNA transcript moves into the next position, leaving the A site open for the entry of another NTP. Only nucleic acids in the active center region (template DNA, turquoise; RNA transcript, red), bridge helix (green), and  $Mg^{2+}$  ions (purple) are shown. The incoming NTP is shown in yellow. (Reprinted by permission of Federation of the European Biochemical Societies from: Boeger, H., Bushnell, D.A., Davis, R. et al. 2005. Structural basis of eukaryotic gene transcription. *FEBS Letters* 579:899–903.)



for bacterial RNA transcription in Section 10.3; however, the details differ because of all the additional factors involved in RNA pol II-mediated transcription.

### Promoter clearance

During abortive initiation, the polymerase synthesizes a series of short transcripts. As RNA pol II moves, it holds the DNA strands apart forming a transcription “bubble.” The upstream edge of the transcription bubble forms 20 bp from the TATA box. TFIIB contacts both DNA strands within the transcription bubble and stabilizes the transcribing complex until a complete 8 or 9 bp DNA–RNA hybrid is formed. The bubble expands downstream until 18 base pairs are unwound and the RNA is at least seven nucleotides long. When this point is reached, the upstream approximately eight bases of the bubble reanneal. This so-called “bubble collapse” marks the end of the need for the TFIIH helicase for transcript elongation. Synthesis of RNA greater than about 10 residues in length leads to displacement of TFIIB from RNA pol II, because the TFIIB-binding site overlaps the RNA exit site (see Fig. 11.9). Promoter clearance requires phosphorylation of the C-terminal domain of RNA pol II at multiple sites within the heptapeptide repeats. There are 27 of these repeats in the yeast RNA pol II CTD. Each repeat contains sites for phosphorylation by specific kinases including one that is a subunit of TFIIH. Hyperphosphorylation of the CTD tail is essential for activating the polymerase and allowing it to begin the elongation phase. During elongation, the polymerase starts to move away from the promoter. Addition of these phosphates helps RNA pol II to leave behind most of the general transcription factors used for initiation. TFIID remains bound and allows the rapid formation of a new preinitiation complex. Once phosphorylated, RNA pol II can unwind DNA, polymerize (synthesize) RNA, and proofread.

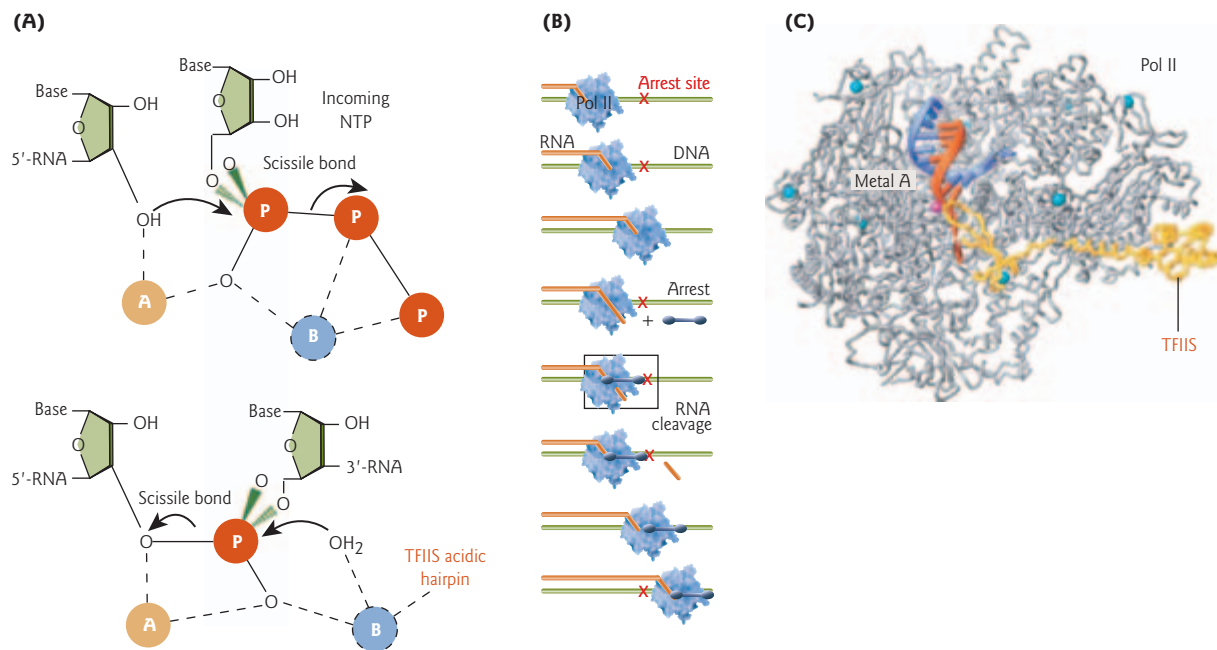
### Elongation: polymerization of RNA

The synthesis of mRNA precursors by eukaryotic RNA pol II is a multistage process consisting of four major steps: initiation, promoter clearance, elongation, and termination. So far, the focus of this chapter has been on events leading up to the initiation of transcription and promoter clearance, and how transcription factors gain access to promoters and enhancers. Although much has been learned about steps involved in initiation, much less is known about later stages in transcription, especially chain elongation. The transition from transcription initiation to elongation is marked by a change in the factors that are associated with RNA pol II.

Transcription elongation is the process by which RNA pol II moves through the coding region of the gene after it initiates mRNA synthesis at the promoter. Incoming (downstream) DNA is unwound before the polymerase active site and is rewound beyond it to form the exiting (upstream) duplex. In the unwound region, the DNA template strand forms a hybrid duplex with growing mRNA. RNA pol II selects NTPs in a template-directed manner. Binding of a NTP to a RNA pol II-transcribing complex with base pairing to the template DNA has been observed by X-ray crystallography. The results are consistent with a four step cycle of nucleotide selection and addition (Fig. 11.26). First, the incoming nucleotide binds to an entry site beneath the active center in an inverted orientation. Second, the NTP rotates into the nucleotide addition site for sampling of correct pairing with the template DNA and for discrimination from dNTPs (that lack the ribose 2′-OH group). Only correctly paired NTPs can transiently bind the insertion site. Third, is the pretranslocation step in which phosphodiester bond formation occurs. Fourth, translocation occurs to repeat the cycle. In the post-translocation complex, the nucleotide just added to the RNA has moved to the next position, leaving the A site open for the entry of a nucleotide. At the upstream end of the hybrid, RNA pol II separates the nascent RNA from the DNA.

### Proofreading and backtracking

In DNA polymerase, the growing DNA shuttles between widely separated active sites for DNA synthesis and cleavage during proofreading (see Fig. 6.13). Surprisingly, in RNA pol II the growing RNA remains at



**Figure 11.27 RNA pol II polymerization and cleavage.** (A) Model of the tunable RNA pol II active site. RNA polymerization and cleavage both require metal ion “A” (e.g.  $Mg^{2+}$ ) in the active center region. The differential positioning of metal ion “B” switches activity from polymerization to cleavage. (Upper panel) Proposed mechanism of nucleotide (NTP) incorporation during RNA polymerization. (Lower panel) Proposed mechanism of TFIIS-mediated RNA cleavage during backtracking. A metal ion “B” interacts either with the phosphates of the incoming NTP (RNA polymerization) or with the TFIIS acidic hairpin and a nucleophilic water molecule (RNA cleavage). (B) A model for the rescue of arrested RNA pol II by TFIIS-mediated mRNA cleavage. When transcribing, RNA pol II (blue) encounters an arrest site (red cross) on DNA (green), it then pauses and backtracks leading to transcriptional arrest. Cleavage of the extruded RNA (orange) is induced by TFIIS (blue). Transcription then continues on past the arrest site. (C) Cut-away view of the crystallographic model of the RNA pol II (gray)–TFIIS (yellow) complex, showing metal ion A in the active site. Template DNA, blue; RNA, red; zinc ions, cyan spheres. (Protein Data Bank, PDB: 1Y1Y. Adapted from Cramer, P. 2004. RNA polymerase II structure. *Current Opinion in Genetics & Development* 14:218–226. Copyright © 2004, with permission from Elsevier.)

a single active site that switches between RNA synthesis and cleavage. X-ray crystallographic analysis of RNA pol II with the elongation factor TFIIS (see below) supports the idea that the polymerase has a “tunable” active site that switches between mRNA synthesis and cleavage (Fig. 11.27). RNA polymerization and cleavage both require a metal ion “A” (e.g.  $Mg^{2+}$ ), but differential coordination of another metal ion “B” switches RNA pol II activity from polymerization to cleavage. In this model, for RNA polymerization, metal B binds the phosphates of the substrate NTP. For cleavage during proofreading, metal B is bound by an additional unpaired nucleotide located in the active center. The tunable polymerase active site allows efficient mRNA proofreading because RNA cleavage creates a new RNA 3′ end at metal A, from which polymerization can continue. Two types of proofreading reactions may occur: removal of a misincorporated nucleotide directly after its addition, or cleavage of a dinucleotide after misincorporation and backtracking by one nucleotide.

During mRNA elongation, RNA pol II also can encounter DNA sequences that cause reverse movement or “backtracking” of the enzyme. During backtracking the register of the RNA–DNA base pairing is maintained but the 3′ end of the transcript is unpaired and extruded from the active center. This can lead to transcriptional arrest. Escape from arrest requires cleavage of the extruded RNA with the help of TFIIS.

For stimulation of the weak nuclease activity of RNA pol II, TFIIS is proposed to insert an acidic  $\beta$ -hairpin loop into the active center to position metal B and a nucleophilic water molecule for RNA cleavage (Fig. 11.27).

## Transcription elongation through the nucleosomal barrier

Although chromatin is remodeled upon transcription initiation, the DNA remains packaged in nucleosomes in the coding region of transcribed genes. As RNA pol II moves through the gene-coding region, the enzyme encounters a nucleosome approximately every 200 bp. How does RNA pol II overcome this nucleosomal barrier? Recent studies suggest the existence of two distinct mechanisms for the progression of RNA polymerases through chromatin: (i) nucleosome mobilization or “octamer transfer” (i.e. movement of the octamer on the DNA); and (ii) H2A-H2B dimer depletion (Fig. 11.28).

### Nucleosome mobilization

Polymerases such as RNA polymerase III and bacteriophage SP6 RNA polymerase appear to use the nucleosome mobilization or “octamer transfer” mechanism for overcoming the nucleosome barrier. In this mode of action, nucleosomes are translocated without release of the core octamer into solution. In other words, DNA is displaced from the nucleosomes and the nucleosomes are transferred to a region of DNA already transcribed by RNA pol II (Fig. 11.28A). This process may be facilitated by the elongation factor FACT (see below). Currently, the mechanism of RNA pol I-dependent chromatin remodeling is unknown.

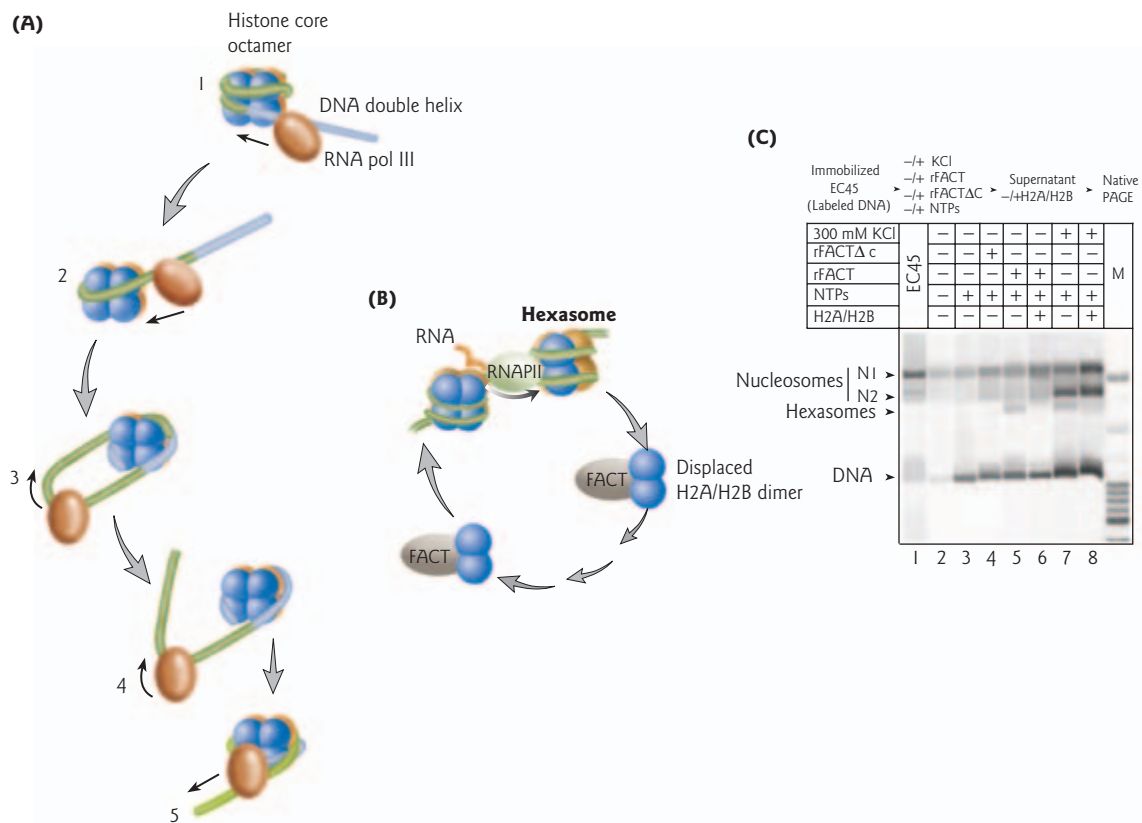
### H2A-H2B dimer removal

Nucleosomes are disrupted on active RNA pol II transcribed genes by H2A-H2B dimer removal (Fig. 11.28B). This mode of overcoming the nucleosome barrier requires a number of auxiliary elongation factors. The first protein factor that has been firmly established to promote RNA pol II elongation *in vitro* is FACT (*facilitates chromatin transcription*). Other complexes that facilitate chromatin transcription *in vitro* are Elongator and TFIIS. Even in the presence of FACT, Elongator, or TFIIS, the rate of RNA pol II transcription through nucleosomes *in vitro* is much slower than the rate of transcript elongation in the cell. *In vitro*, a significant number of RNA pol II molecules are blocked by the nucleosome, suggesting that overcoming the nucleosome barrier requires the coordinated action of additional, unidentified factors.

### FACT promotes nucleosome displacement

FACT plays a role in the elongation of transcripts through nucleosome arrays by promoting transcription-dependent nucleosome alterations. FACT is composed of two protein subunits, designated hSpt16 and SSRP1 in humans. Spt6 was first discovered by a genetic screen in yeast called SPT (suppressor of TY) performed more than two decades ago. This screen uncovered transcription factor mutations that overcome the consequences of inserting a foreign piece of DNA (a TY transposon element) in the promoter of a reporter gene. The FACT complex has a DNA-binding domain and a domain that interacts with histones H2A and H2B. The large subunit (hSpt16) also interacts with the catalytic subunit of a histone acetyltransferase (HAT). This interaction provides a possible mechanism explaining the extent of acetylation observed at transcriptionally active regions. A model for FACT's mode of action has been developed, based on an *in vitro* assay system that allows the study of transcription through positioned nucleosomes. In this model, FACT enables the displacement of a dimer of H2A-H2B in front of RNA pol II, leaving a histone “hexamer” at the same location as the initial octamer. FACT does not require ATP hydrolysis for its mode of action. After passage of the polymerase, FACT enables the immediate reassembly of the H2A-H2B dimer (Fig. 11.28B).





**Figure 11.28 Mechanisms of transcription through the nucleosome by RNA polymerases.** (A)

Nucleosome mobilization or “octamer transfer” model for RNA pol III. (1) RNA pol III rapidly transcribes the first ~25 bp of nucleosomal DNA causing (2) partial dissociation of the DNA from the core octamer (blue). DNA that has been transcribed is shown in blue. DNA that has not been transcribed is green. (3) The DNA behind RNA pol III transiently binds to the exposed surface of the octamer forming a loop. When the polymerase has moved ~60 bp into the core, the downstream portion of the DNA dissociates from the octamer (4), completing octamer transfer. (5) Transcription proceeds to the end of the template. (Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., Felsenfeld, G. 2004. Chromatin remodeling by RNA polymerases. *Trends in Biochemical Sciences* 29:127–135.)

(B) Histone H2A-H2B dimer depletion model for RNA pol II. FACT mediates displacement of the H2A-H2B dimer from the core octamer, leaving a “hexosome” on the DNA. The histone chaperone activity of FACT helps to redeposit the dimer on the DNA after passage of RNA pol II (RNAPII). (C) Experiment demonstrating that FACT-facilitated transcription through the nucleosome results in the formation of hexasomes. Immobilized nucleosomes containing end-labeled DNA were transcribed *in vitro* in the presence or absence of NTPs, excess H2A-H2B dimer, recombinant FACT (rFACT), a nonfunctional mutant lacking the C-terminus (rFACTΔC), or 300 mM KCl as a control. This control was used as it had been previously shown that nucleosomes transcribed in high salt release the H2A-H2B dimer, forming a hexosome. Transcribed nucleosomes were analyzed on a native gel. The positions of polymerase-free nucleosomes N1 and N2, hexasomes, and DNA are indicated by arrowheads. M, molecular weight markers. (Reprinted with permission from: Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., Reinberg, D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science* 301:1090–1093. Copyright © 2003 AAAS.)



### Elongator facilitates transcript elongation

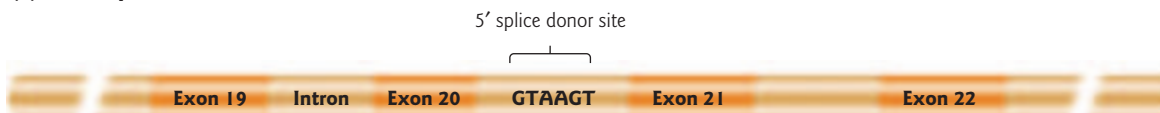
The Elongator complex was first isolated in yeast, but its exact function in transcript elongation remains elusive (Disease box 11.4). Elongator is composed of six major subunits, designated ELP1 to ELP6. ELP3 has been shown to have HAT activity. Subsequently, the human Elongator complex was purified from a human



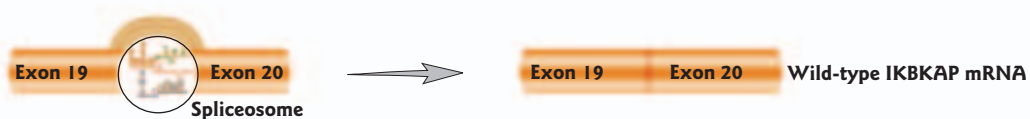
## DISEASE BOX 11.4

## Defects in Elongator and familial dysautonomia

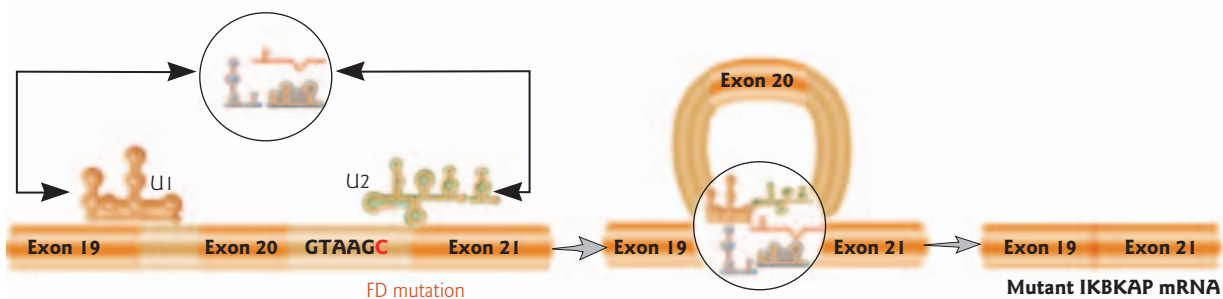
(A)

(a) *IKBKAP* pre-mRNA

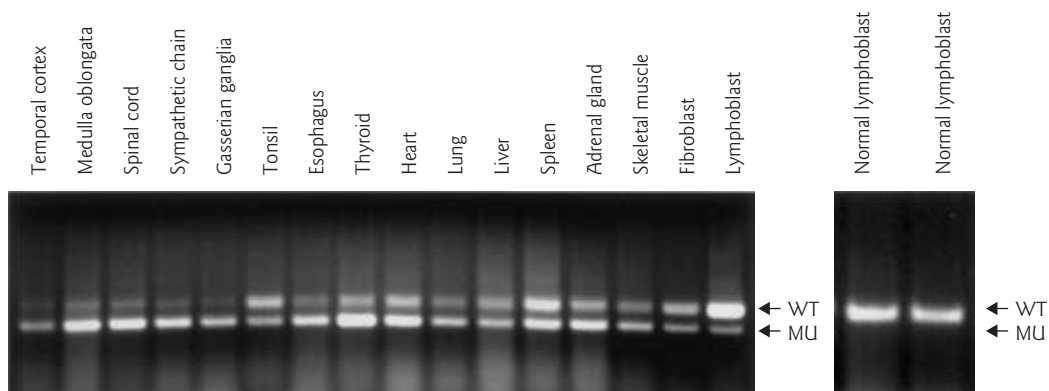
(b) Normal splicing



(c) FD mis-splicing



(B)





Familial dysautonomia is a disorder of the sensory and autonomic nervous system, which involves progressive depletion of unmyelinated sensory and autonomic neurons. It is inherited as an autosomal recessive and is very common in the Ashkenazi Jewish population; there is a carrier frequency of one in 30 descendents of Jews from Germany, Poland, and Austria, and Eastern Europe.

### Symptoms of familial dysautonomia

The autonomic nervous system controls such involuntary functions as swallowing, digestion, and regulation of body temperature and blood pressure. The most distinctive feature of the disorder is the absence of tears when crying. The loss of neuronal function also results in decreased perception of heat, pain, and taste. For example, a person with familial dysautonomia leaning on a pot of boiling water may not feel it and could be seriously burned. Other symptoms include breath-holding episodes, vomiting in response to stress, profuse sweating, and spinal curvature (scoliosis) in 90% of patients by age 13. The average life expectancy is approximately 30 years, and patients frequently have long hospital stays. Treatment currently is mainly preventative and supportive; for example, use of artificial tears, special feeding therapies, and protection from injury.

### A defect in the *IKBKAP* gene causes familial dysautonomia

Two mutations in the gene *IKBKAP* are responsible for familial dysautonomia. *IKBKAP* was originally identified as the gene encoding a protein that was thought to function as an I- $\kappa$ B kinase-associated protein (see Section 11.10), hence the name “IKAP.” Subsequent reports determined that, in fact, IKAP is a subunit of the Elongator complex, homologous to the yeast ELP1 protein.

Neither of the two mutations in *IKBKAP* result in failure to express the IKAP protein. The most common mutation (> 99.5% of cases) is a splice site mutation that results in tissue-specific skipping of one exon during mRNA splicing. The *IKBKAP* gene contains 37 exons. A single T  $\rightarrow$  C change at base pair 6 of the intron 20 donor splice site leads to symptoms in homozygotes. There is a decrease in splicing efficiency with sporadic skipping of exon 20, reducing the level of wild-type *IKBKAP* mRNA. Every familial dysautonomia cell type expresses both wild-type and mutant IKAP, the ratio of which is variable (Fig. 1). *IKBKAP* mRNA is primarily wild type in fibroblasts and in lymphoblast cell lines. However, in the brain, mutant mRNA with exon 20 missing is primarily expressed. It is not clear why brain cells are particularly sensitive to exon skipping. A second extremely rare mutation is a single G  $\rightarrow$  C change in an exon, which causes an arginine to proline missense mutation that disrupts a threonine phosphorylation site in the protein. This mutation has only been seen in heterozygous patients and has never been detected in the homozygous state.

Researchers are still puzzling over whether a defect in transcription elongation leads to the degeneration of specific sensory and autonomic neurons in humans. Alternatively, the disease symptoms could be caused by disruption of an as yet unidentified cellular pathway in which IKAP alone plays a role. There is some evidence that IKAP may have multiple roles in the cell. A recent study suggests that ELP1P, the yeast homolog of IKAP, negatively regulates polarized exocytosis of post-Golgi secretory vesicles. Since neurons are highly dependent on polarized exocytosis for their development and function, this could make them especially sensitive to dysregulation of IKAP function.

**Figure 1 (opposite) Familial dysautonomia.** (A) Schematic representation of the aberrant splicing of *IKBKAP* pre-mRNA seen in familial dysautonomia (FD). The wild-type splice site sequence (GTAAGT) in between exons 20 and 21 of *IKBKAP* pre-mRNA is shown above. In normal splicing, the 5' splice donor site is recognized by the spliceosome (see Section 13.5), resulting in joining of exons 19 and 20. In FD missplicing, the major FD mutation at base pair 6 of intron 20 is shown in red. This mutation decreases the efficiency of splicing and sometimes results in the skipping of exon 20. (B) Expression of wild-type (WT) and mutant (MU) *IKBKAP* mRNA in postmortem FD tissue samples and cell lines. All samples were assayed by reverse transcription–polymerase chain reaction (RT-PCR) and PCR products were fractionated by electrophoresis on an agarose gel stained with ethidium bromide (see Tool box 8.6 and Fig. 9.8D for methods). The relative WT : MU transcript ratio was consistently observed to be highest in FD lymphoblast lines, lowest in FD nervous system tissues, and intermediate in other tissues. Mutant *IKBKAP* transcript is never detected in RNA from normal cells or tissues. (Reprinted from Cuajungco, M.P., Leyne, M., Mull, J. et al. 2003. Tissue-specific reduction in splicing efficiency of *IKBKAP* due to the major mutation associated with familial dysautonomia. *American Journal of Human Genetics* 72:749–758. Copyright © 2003 by The American Society of Human Genetics, with permission of the University of Chicago Press.)

cell line (HeLa cells) by column chromatography. Like the yeast complex, human Elongator is also composed of six subunits, including a HAT with specificity for histone H3 and to a lesser extent for histone H4. Elongator directly interacts with RNA pol II and facilitates transcription, but it does not appear to interact directly with any other elongation factors, including FACT.

### **TFIIS relieves transcriptional arrest**

The elongation factor, TFIIS, facilitates passage of RNA pol II through regions of DNA that can cause transcription arrest. These sites include AT-rich sequences, DNA-binding proteins, or lesions in the transcribed DNA strand. TFIIS rescues transcriptional arrest by a backtracking mechanism that stimulates endonucleolytic cleavage of the nascent RNA by the RNA pol II active center (see Fig. 11.27).

## **11.9 Nuclear import and export of proteins**

One of the hallmarks of eukaryotic cells is the compartmentalization of the genome into a separate organelle called the nucleus. The sections in this chapter so far have focused on proteins involved in regulating gene transcription, including transcription factors, the general transcription machinery, coactivators, and corepressors. Since protein synthesis occurs in the cytoplasm, this means that transcriptional regulatory proteins must be delivered to their site of activity in the nucleus. The control of nuclear localization of transcriptional regulatory proteins represents a level of transcriptional regulation in eukaryotes that is not present in prokaryotes.

Trafficking between the nucleus and the cytoplasm occurs via the nuclear pore complexes (NPCs). NPCs are large multiprotein complexes embedded in the nuclear envelope – the double membrane that surrounds the nucleus (Focus box 11.6). The NPCs allow bidirectional passive diffusion of ions and small molecules. In contrast, nuclear proteins, RNAs, and ribonucleoprotein (RNP) particles larger than ~9 nm in diameter (and greater than ~40–60 kD) selectively and actively enter and exit the nucleus by a signal-mediated and energy-dependent mechanism (Fig. 11.29). Proteins are targeted to the nucleus by a specific amino acid sequence called a nuclear localization sequence (NLS). In some cases, a nuclear protein without a NLS dimerizes with an NLS-bearing protein and rides “piggyback” into the nucleus. In addition, some nuclear proteins shuttle repeatedly between the nucleus and cytoplasm. Their exit from the nucleus requires a nuclear export sequence (NES). Nuclear import and export pathways are mediated by a family of soluble receptors referred to as importins or exportins, and collectively called karyopherins (Table 11.3). The presence of several different NLSs and NESs and multiple karyopherins suggests the existence of multiple pathways for nuclear localization.

### **Karyopherins**

Karyopherins are proteins composed of helical molecular motifs called HEAT repeats (the acronym is derived from the four name-giving proteins: *Huntington*, *Elongation factor 3*, the “*A*” subunit of protein phosphatase 2, and *TOR1* kinase) or Armadillo repeats (so named for their discovery in the *Drosophila* Armadillo protein). The repeats are stacked on top of each other to form highly flexible superhelical or “snail-like” structures (Fig. 11.29). The largest class of soluble receptors is the karyopherin- $\beta$  family, which is involved in the transport of proteins and RNP cargoes (Table 11.3). A second group of receptors, which is structurally unrelated to the karyopherin- $\beta$  family, is the family of nuclear export factors (NXFs) that are involved in the export of many mRNAs. A third class is represented by the small nuclear transport factor 2 (NTF2), which imports the small GTPase Ran into the nucleus.

Of the more than 20 members of the karyopherin- $\beta$  family in vertebrates, 10 of these play a role in nuclear import, whereas seven function in nuclear export (Table 11.3). Importin- $\beta 1$  is one of the predominant karyopherins that drives import. Although a small number of cargo proteins may bind importin- $\beta 1$  directly, most cargoes require the adaptor protein importin- $\alpha$ . Seven importin- $\alpha$  adapters have

**Table 11.3 Nuclear import and export factors of the karyopherin- $\beta$  family.**

Name	Cargo
<b>Import receptors</b>	
Importin- $\beta$ 1	Many cargoes with basic NLSs via an importin- $\alpha$ adapter, snRNPs via snurportin
Karyopherin- $\beta$ 2	mRNA-binding proteins, histones, ribosomal proteins
Transportin SR	mRNA-binding (phospho-SR domain) proteins
Transportin SR2	HuR
Importin 4	Histones, ribosomal proteins
Importin 5	Histones, ribosomal proteins
Importin 7	Glucocorticoid receptor, ribosomal proteins
Importin 8	SRP19
Importin 9	Histones ribosomal proteins
Importin 11	UbcM2, ribosomal protein L12
<b>Export receptors</b>	
Crm1	Leucine-rich NES cargoes
Exportin-t	tRNA
CAS	Karyopherin $\alpha$
Exportin 4	eIF-5A
Exportin 5	microRNA precursors
Exportin 6	Profilin, actin
Exportin 7	p50Rho-GAP, 14-3-3 $\delta$
<b>Import/export</b>	
Importin 13	Rbm8, Ubc9, Pax6 (import) eIF-1A (export)

NES, nuclear export sequence; NLS, nuclear localization sequence.

been characterized in mammals (importin  $\alpha$ 1– $\alpha$ 7). Although interchangeable for many cargoes *in vitro*, there are reports of preferential use of specific importin- $\alpha$  adapters *in vivo*. These receptors bind to signals in their cargo and to a subset of nucleoporins containing repeats of the amino acids phenylalanine and glycine (designated as FG repeats). Binding to FG repeats mediates passage through the NPCs. There is some redundancy between certain transport pathways. For example, *in vitro* import assays in mammalian cells have shown that five different importins can mediate nuclear entry of histones, and at least four importin- $\beta$ -like factors are able to transport ribosomal proteins into the nucleus for assembly into ribosomes.

### Nuclear localization sequences (NLSs)

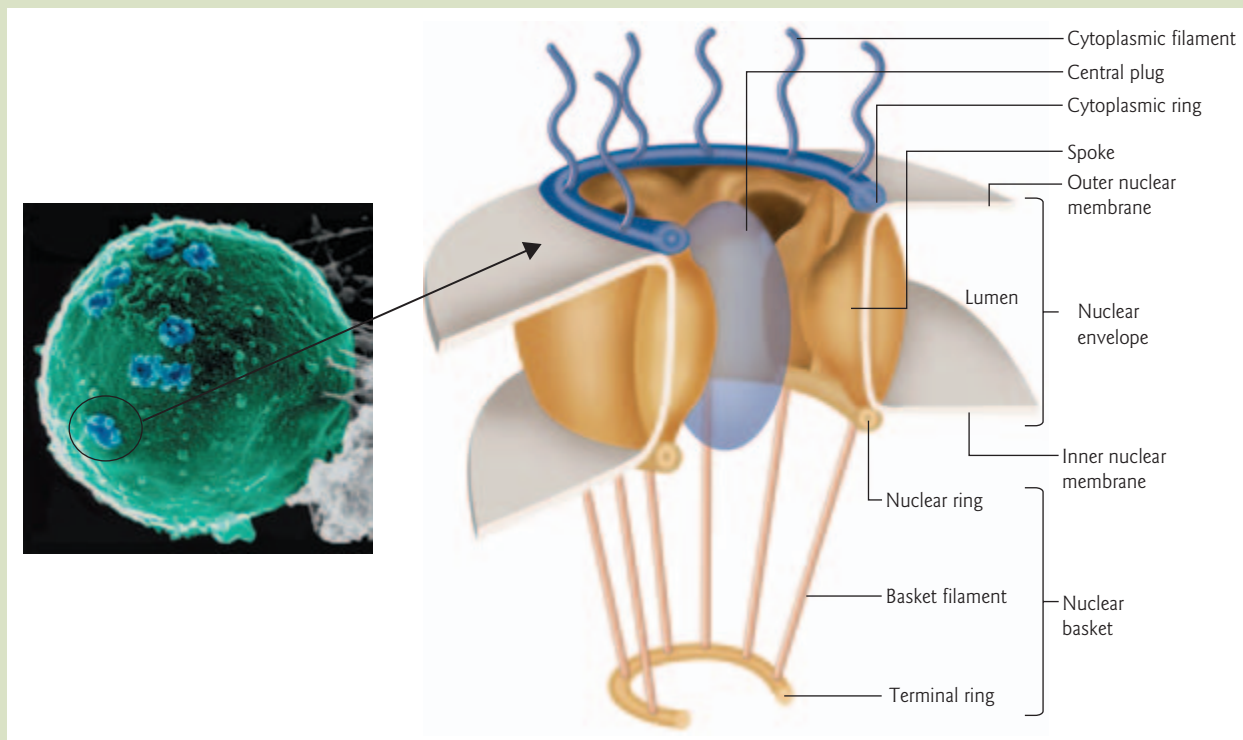
Unlike signal sequences targeting proteins to the endoplasmic reticulum or mitochondrion, which are generally removed from proteins during transit, nuclear proteins retain their NLS. This may be to ensure



Over the past few years there has been much progress in knowledge of the structure and molecular architecture of the nuclear pore complexes (NPCs). NPCs are highly dynamic, modular machines embedded in the nuclear membrane or “nuclear envelope” (Fig. 1). The nuclear envelope that surrounds the nucleus is a double lipid bilayer. There is an outer membrane, a lumen (perinuclear space of 20–40 nm), and an inner membrane. Electron microscopy thin sections show that the outer membrane is continuous with the membrane of the endoplasmic reticulum, and its outer surface is studded with ribosomes like the cytoplasmic face of the rough endoplasmic reticulum. The NPCs serve as a gateway for the exchange of material between the nucleus and cytoplasm. The nucleus of a human cell has several thousand NPCs connected by the nuclear lamina. The whole NPC structure is ~50 MDa in

size in yeast. In mammals, estimates range from 60 to 125 MDa. The overall architecture appears to be well conserved between species despite differences in mass.

The most striking feature of the NPCs is their eight-fold radial symmetry. They are composed of eight globular subunits that form a central spoke–ring complex, including a cylindrical structure which surrounds the central translocation channel (Fig. 1). Ring-like structures flank the spoke–ring complex on both its cytoplasmic and nuclear side. Eight fibrils are attached to each of these rings. On the cytoplasmic face of the NPC the fibrils have free ends that extend into the cytoplasm. On the nucleoplasmic face the fibrils form a basket-like structure that ends in a terminal ring. The NPCs are composed of multiple copies of a set of proteins that are collectively referred to as nucleoporins. Proteomic analysis of NPCs from both yeast and mammals



**Figure 1 The nuclear pore complex (NPC).** Field emission scanning electron microscope image of the yeast nucleus. Blue pseudocoloring highlights the NPCs; green pseudocoloring highlights the nuclear envelope together with attached ribosomes. Scale bar, 100 nm. (Reprinted by permission of Nature Publishing Group and Macmillan Publishers Ltd: Kiseleva, E. 2004. Cell of the month. *Nature Cell Biology* 6:497. Copyright © 2004. Photograph kindly provided by Elena Kiseleva, Institute of Cytology and Genetics, Novosibirsk-90, Russia). (Inset) A cut-away model of the NPC showing the key structural features. The NPC consists of eight spokes that form a cylinder embedded in the nuclear envelope. Several ring-like structures connect the spokes. Additional filamentous structures point towards the lumen of the nucleus and cytoplasm.

# The nuclear pore complex

FOCUS BOX 11.6



has shown that there are about 30 different nucleoporins. Because of the eight-fold symmetry of the NPC, each nucleoporin is present at a copy number of either eight or an integer multiple of eight. Individual nucleoporins have unique roles in regulating NPC function and nuclear import and export. Immunogold electron microscopy has shown that most of the nucleoporins are located symmetrically on both faces of the NPC, whereas only a few are located asymmetrically on either the nuclear face or the cytoplasmic face.

The central channel in the NPC connecting the cytoplasm to the nucleus has a length of about 90 nm. The nuclear

basket and the central channel appear to dilate in response to the translocation of large cargoes, increasing the resting 9–11 nm channel to an effective diameter of ~45–50 nm. When viewed in projection, the central pore often appears to be obstructed by a particle that varies greatly in size and shape and is commonly referred to as the “central plug” or “transporter” (Fig. 1). Three-dimensional reconstruction of native NPC and several studies using atomic force microscopy suggest that the central plug represents a view of the distal ring of the nuclear basket and/or cargo translocating through the central pore of the NPC.

that they can reaccumulate in the nucleus after each mitotic cell division. The best studied NLSs are basic amino acid sequences, typically rich in lysine and arginine. Despite the impressive number of receptor–cargo interactions that have been studied, the prediction of NLSs in candidate proteins remains extremely difficult. There is no real consensus sequence for an NLS. Some proteins, such as nucleoplasmin, have a bipartite NLS (Table 11.4, Focus box 11.7). NLSs interact with members of the karyopherin  $\beta$  family either directly or through the adapter importin- $\alpha$ .

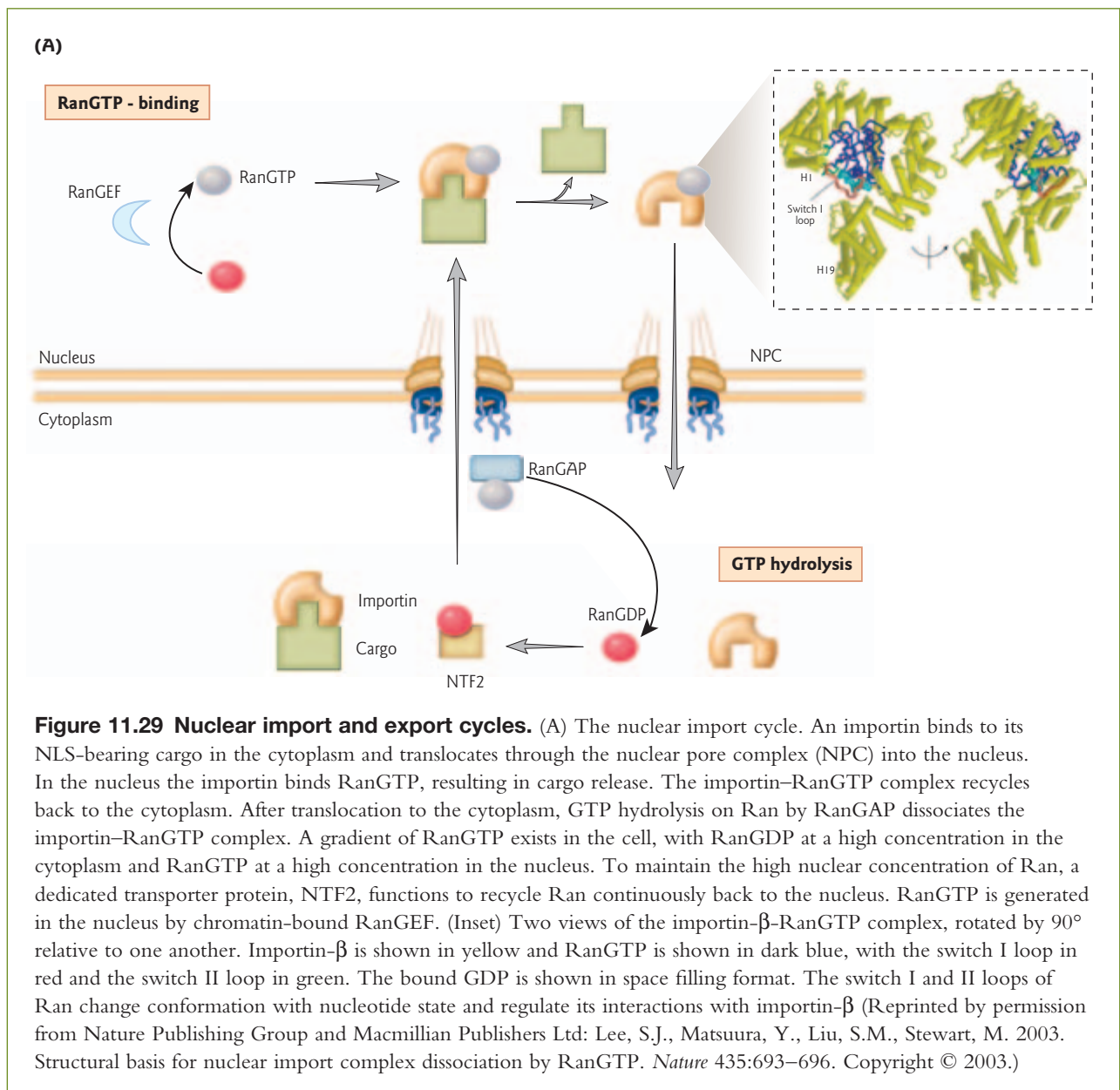
## Nuclear export sequences (NESs)

Of the nuclear export signals found in eukaryotes, the best characterized are the small hydrophobic leucine-rich NESs, first described in the human immunodeficiency virus 1 (HIV-1) Rev protein (Table 11.4). These classic Rev-type NESs function via interaction with the export factor CRM1. CRM1 was named for a mutant of the fission yeast *Schizosaccharomyces pombe* that was shown to be essential for proliferation and for “chromosome region maintenance.” Although nuclear transport of RNP complexes by members of the

**Table 11.4 Nuclear localization sequences and nuclear export sequences.**

Signal type	Protein	Amino acid sequence*
Nuclear localization sequence (NLS)	SV-40 T antigen (monopartite NLS)	PKKKRKV
	Nucleoplasmin (bipartite NLS)	KRPAATKKAGQAKKKKLD
Nuclear export sequence (NES)	HIV-1 Rev	L-PPL-ERLTL
	I $\kappa$ B $\alpha$	MVKEL-QEIRL
	TFIIIA	L-PVL-ENLTL
	Consensus	$\phi$ <sub>x</sub> <sub>2-3</sub> $\phi$ <sub>x</sub> <sub>2-3</sub> $\phi$ <sub>x</sub> $\phi$ ( $\phi$ = conserved hydrophobic residues L, I, V, F, or M; x is any amino acid)

\* See Fig. 5.3 for standard one-letter amino acid designations.



karyopherin- $\beta$  family is often specified by the signals present on their protein components, two exportins, exportin-t and exportin 5, interact with tRNA directly (see Table 11.3).

### Nuclear import pathway

Factors involved in nuclear import have been well studied biochemically. Recent advances into regulatory mechanisms have been made by combining computer simulation and real-time assays to test the predictions in intact cells. The process of nuclear import involves three main steps: (i) cargo recognition and docking; (ii) translocation through the nuclear pore complex (NPC); and (iii) cargo release and receptor recycling (see Fig. 11.29A).



**Figure 11.29 (cont'd)** (B) Nuclear export cycle. Exportins bind to their cargo in the nucleus in the presence of RanGTP. In the cytoplasm, GTP hydrolysis causes disassembly of the export complex and recycling of the export receptor. (Inset) Two orthogonal views showing the structure of an export complex containing Cse1p (yellow), Kap60p (importin- $\alpha$ ), and RanGTP (blue). Cse1p envelopes RanGTP and the C-terminal region of importin- $\alpha$  (green) and its IBB (importin- $\beta$ -binding) domain. The “spring-loaded” complex disassembles spontaneously upon RanGTP hydrolysis. The Cse1p  $\alpha$ -helices are represented by cylinders. The ARM repeats of Kap60p and HEAT repeats of Cse1p are labeled A1–A10 and H1–H20, respectively. GTP is shown as space-filling spheres. (Reprinted by permission from Nature Publishing Group and Macmillan Publishers Ltd: Matsuura, Y., Stewart, M. 2004. Structural basis for the assembly of a nuclear export complex. *Nature* 432:872–877. Copyright © 2004.)

## Cargo recognition and docking

The import receptor for the classic, lysine/arginine-rich NLS is a complex of an importin- $\alpha$  adapter and importin- $\beta$ 1. Members of the importin- $\alpha$  family contain a C-terminal region that binds directly to cargo proteins containing an NLS. Importin- $\beta$ 1 binds to both importin- $\alpha$  and to nuclear pore complex (NPC) proteins. Binding occurs at the cytoplasmic filaments of the NPC. Cargo recognition and docking does not require energy from either ATP or GTP hydrolysis.

## Translocation through the nuclear pore complex

The exact molecular mechanism by which cargo translocation through the NPC occurs is poorly understood. However, weak hydrophobic interactions between importins and the FG repeat domains of nucleoporins seem to be essential. For many years it was assumed that translocation through the NPC would occur by some type of ATP-driven motor. However, when it became evident through experimentation that neither ATP nor GTP hydrolysis were required for translocation, different models had to be developed. An older “affinity gradient” model suggested that karyopherins travel along a gradient of nucleoporin-binding



## FOCUS BOX 11.7

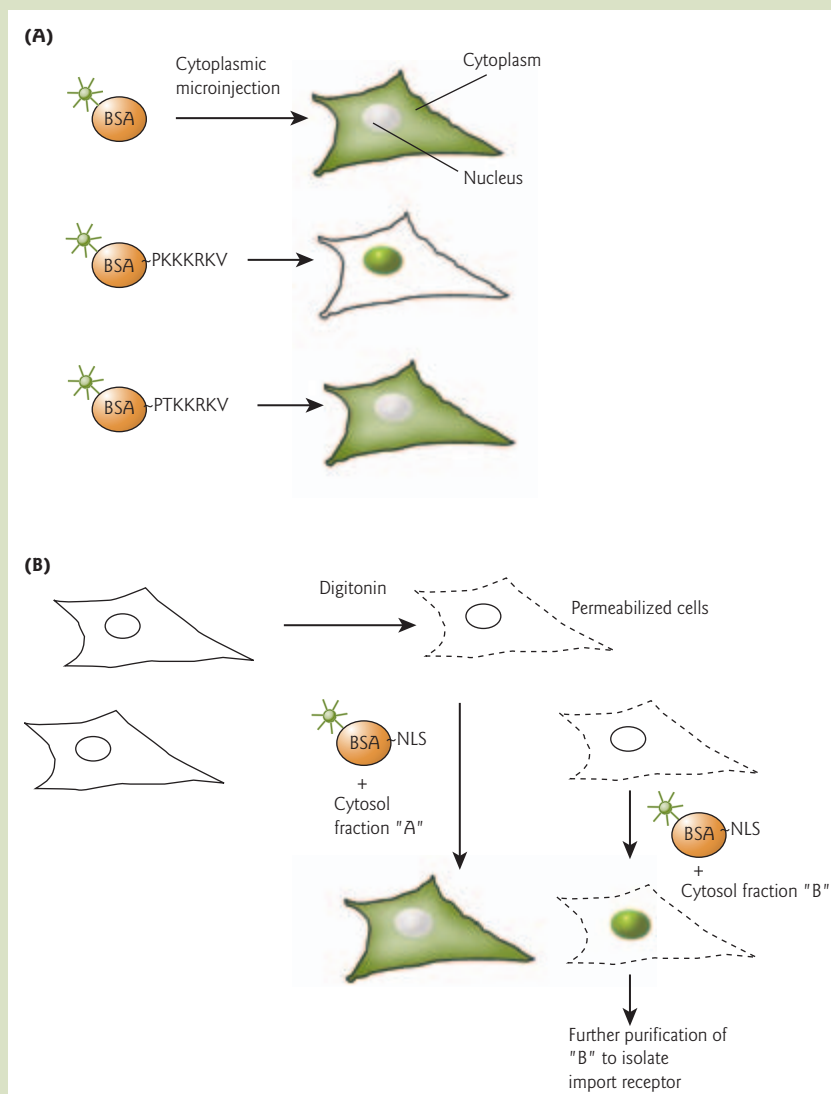
## Characterization of the first nuclear localization sequence

The first nuclear localization sequence (NLS) was characterized in 1984 for the simian virus 40 (SV40) T antigen (the helicase that initiates DNA replication, see Fig. 6.7). The > 60 kDa antigen accumulates within nuclei of infected mammalian cells. Daniel Kalderon and co-workers showed that when the lysine at position 128 in the polypeptide chain (Lys128) was mutated to threonine (Thr),

the T antigen was unable to enter the nucleus. Subsequent analysis revealed that mutation of neighboring amino acids also affected nuclear import. This observation led to the definition of a positively charged region around Lys128 that seemed responsible for nuclear entry. The region, termed the classic NLS, included the amino acid sequence: Pro-Lys-Lys128-Lys-Arg-Lys-Val (or PKKKRKV).

### Figure 1 Characterization of nuclear localization sequences and import receptors.

(A) When bovine serum albumin (BSA) tagged with fluorescein (green) was microinjected into the cytoplasm of cells, as expected the serum protein did not enter the nucleus. Attaching the short amino acid sequence “PKKKRKV” from the SV40 T antigen to the fluorescein-tagged BSA allowed nuclear entry. When threonine was substituted for one of the lysines (PTKKRKV), BSA was unable to enter the nucleus. These experiments demonstrated that the SV40 T antigen nuclear localization sequence (NLS) was necessary and sufficient for nuclear import. (Source: Kalderon D., Richardson W.D., Markham, A.F., Smith, A.E. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* 311:33–38.) (B) Biochemical fractionation of cytosol led to the discovery of the various soluble receptors that bind specifically to NLS-containing. Permeabilized cells were incubated with fluorescein-tagged BSA-NLS and cytosol fractions, and then assayed for nuclear entry of the protein by fluorescence microscopy.





## Characterization of the first nuclear localization sequence

FOCUS BOX 11.7



Decisive evidence that the classic NLS sequence was necessary and sufficient for nuclear entry was provided as follows. When bovine serum albumin (BSA) tagged with fluorescein was microinjected into the cytoplasm, as expected the serum protein did not enter the nucleus (Fig. 1). However, attaching the short polypeptide sequence “PKKKRKV” to the fluorescein-tagged BSA (fluorescein-BSA:NLS) allowed nuclear entry. Injecting more of the fluorescein-BSA:NLS soon saturated the pathway. These early experiments predicted that NLSs within nuclear proteins bind to saturable receptors at the NPCs. Thus, molecular biologists began to search for such receptors.

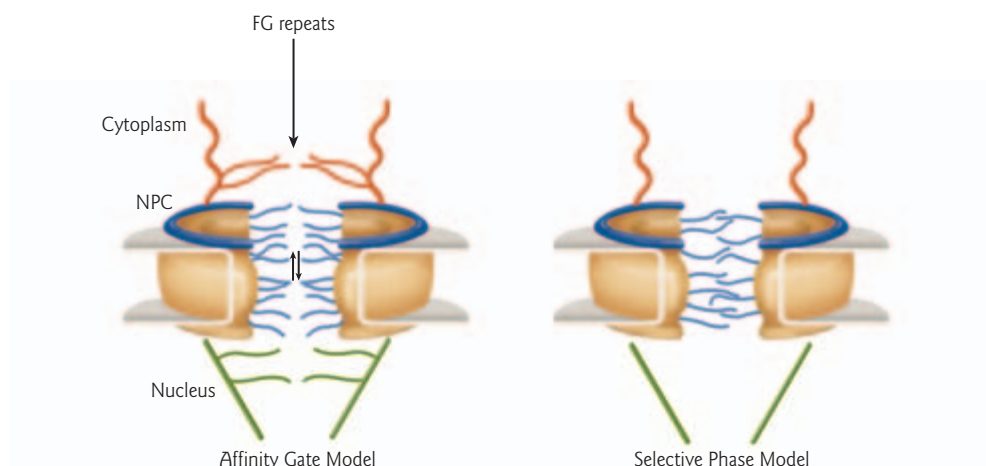
An *in vitro* assay was developed, in which the plasma membrane of mammalian cells is permeabilized by digitonin, a weak detergent that leaves the nuclear membrane intact. Cells were then incubated with fluorescein-BSA:NLS. Nuclear entry of fluorescein-BSA:NLS was monitored by fluorescence microscopy. Entry in buffer alone was inefficient but could be greatly enhanced by adding “cytosol” (contents of cytoplasm, excluding the various membrane-bound organelles) from another cellular source. Biochemical fractionation of the cytosol led to the discovery of the various soluble receptors that bind specifically to NLS-containing proteins.

sites, encountering nucleoporins of increasing affinity during their translocation. However, the cytoplasmic filaments of the NPC can be removed with no obvious effect on transport. Also, experiments have shown that the direction of transport can be inverted in the absence of the RanGTP gradient, and deletion of all asymmetric FG repeats has only modest effects on the nuclear import pathway. These findings suggest that the NPC has no intrinsic directionality. The two most widely supported models are the Brownian affinity or “virtual gating” model and the selective phase or “sieve” model (Fig. 11.30). Both of these models (and other variations on the general theme) agree that transport through the central channel occurs via diffusion, but differ in the basis for the permeability barrier in the NPC and the structure of the central channel.

Nucleocytoplasmic transport through the NPC occurs against a concentration gradient, and because this cycle is bidirectional, an energy source and a directional cue are needed. Both are provided by the small GTPase Ran (see Fig. 11.29). The 24 kDa protein Ran (for *Ras*-related nuclear) belongs to a superfamily of GTP-binding proteins that act as molecular switches cycling between GDP- and GTP-bound states. The conversion from the GDP- to GTP-bound state involves nucleotide exchange. In contrast, the conversion of RanGTP to RanGDP occurs by removal of the terminal phosphate from the bound GTP. RanGTP is in a high concentration within the nucleus and a low concentration in the cytoplasm. The concentration difference between free nuclear and cytoplasmic RanGTP is estimated to be at least 200-fold. This gradient is created by an asymmetric distribution of the Ran guanosine-nucleotide exchange factor (RanGEF; also known as RCC1) and the Ran-specific GTPase-activating protein (RanGAP). RanGEF is a resident nuclear protein that binds to nucleosomes through an interaction with histones H2A and H2B and promotes nucleotide exchange, replacing the GDP bound to Ran with GTP. RanGAP, on the other hand, is excluded from the nucleus and acts to maintain Ran in the GDP-bound state in the cytoplasm. RanGAP is found at its highest concentration at the outer face of the NPC where it associates with the nucleoporin RanBP2 via its SUMO modification. The actual process of NPC translocation itself does not require RanGTP hydrolysis. None of the ~30 proteins in the NPC proteome encodes motor or ATPase domains, providing further evidence that the complex does not actively pump cargoes across the nuclear membrane.

### Cargo release and receptor recycling

Once the cargo–import receptor complex reaches the nuclear side of the NPC, RanGTP binds to the importin and dislodges it from the cargo (see Fig. 11.29A). Cargo dissociation occurs by an allosteric



**Figure 11.30 Schematic illustrations of the nuclear pore complex (NPC) translocation models.** Two models are depicted for the selectivity of translocation through the NPC. (Left) In the affinity gate model the FG repeat-containing nucleoporins form multiple waving filaments at both ends of the NPC. The FG repeats function as a selective barrier to diffusion of non-nuclear proteins. The affinity of importins or exportins for these flexible filaments, moved by Brownian motion, increases the residence time of karyopherin–cargo complexes at the nuclear periphery and the probability of their access to the NPC central channel. Proteins gain access to the channel and diffuse through the NPC by “hopping” of the karyopherin–cargo complex from FG repeat to FG repeat. High-affinity nucleoporin-binding sites present in the destination compartment support the directionality of transport. (Right) As in the affinity gate model, the selective phase model assumes that the karyopherin–cargo complex moves randomly by diffusion through the central channel. The hydrophobic FG repeats are proposed to form a selective semiliquid phase or meshwork into which the karyopherins partition. These interactions enable exclusion of non-nuclear proteins that cannot dissolve into the sieve. The hydrophobic karyopherins in association with cargo enter (dissolve into) the sieve structure and cross the NPC, via their transient and low-affinity interactions with FG repeats.

mechanism in which binding of the importin to RanGTP results in a conformational change that is transmitted to its cargo-binding domain. At this point in the pathway, the cargo is free to carry out its nuclear function. The RanGTP–importin complex subsequently translocates back to the cytoplasm. For the importin- $\alpha$ /importin- $\beta$ 1-mediated pathway, the exportin CAS (also known as Cse1p) binds to importin- $\alpha$  and mediates its nuclear export in association with RanGTP.

The receptor recycling step is the only energy-requiring step in nuclear import – at least two GTP molecules are consumed per NLS import cycle. GTP hydrolysis occurs with the aid of the accessory proteins RanGAP and RanBP1 converting RanGTP to RanGDP. After GTP hydrolysis, the export complexes dissociate and the importins are recycled for another round of import (see Fig. 11.29A). RanGDP is rapidly imported into the nucleus by transport factor NTF2, where it is converted to RanGTP with the aid of RanGEF. Several million molecules of Ran have to be imported every minute into the nucleus of an actively dividing mammalian cell to keep up with the demands of nuclear import and export cycles.

### Nuclear export pathway

As described above, RanGTP has very different functions in import and export. In the nuclear import pathway it causes disassembly of import complexes. But, in nuclear export it is required for the assembly of a cargo–exportin complex. The cargo may be importin- $\alpha$ , or a nuclear protein that shuttles between the nucleus and cytoplasm. Upon binding, RanGTP induces a conformational change in the export receptor

that results in energy storage. The exportin is twisted from an S shape into a horseshoe-like conformation in which both arches of the “horseshoe” interact with RanGTP. When the export complex translocates through the central channel of the NPC and arrives on the outer (cytoplasmic) face it is exposed to proteins that cause GTP hydrolysis (see above). This releases the energy stored in the “spring-loaded” exportin which “pops” open and releases its cargo (see Fig. 11.29B).

## 11.10 Regulated nuclear import and signal transduction pathways

Any protein with both an NLS and NES has the potential to shuttle back and forth between the nucleus and cytoplasm. If such a protein is a transcription factor, this has clear implications for post-translational regulation. For example, a transcription factor may be sequestered in the cytoplasm at a particular developmental stage or in unstimulated cells, and may remain cytoplasmic until an extracellular signal induces its nuclear import. Spatial separation of the transcription factor from its DNA target (by exclusion from the nucleus) acts as a potent inhibitor of function. Regulation of NLS and NES activity can occur by several mechanisms, including post-translational modifications (e.g. phosphorylation and dephosphorylation) that mask (or unmask) the NLS or NES. Where a protein is localized at “steady state” depends on the balance between import, retention, and export, and which signal is dominant. Nuclear retention may be mediated by domains of the protein that interact with components of the nucleus, such as chromatin or the nuclear matrix.

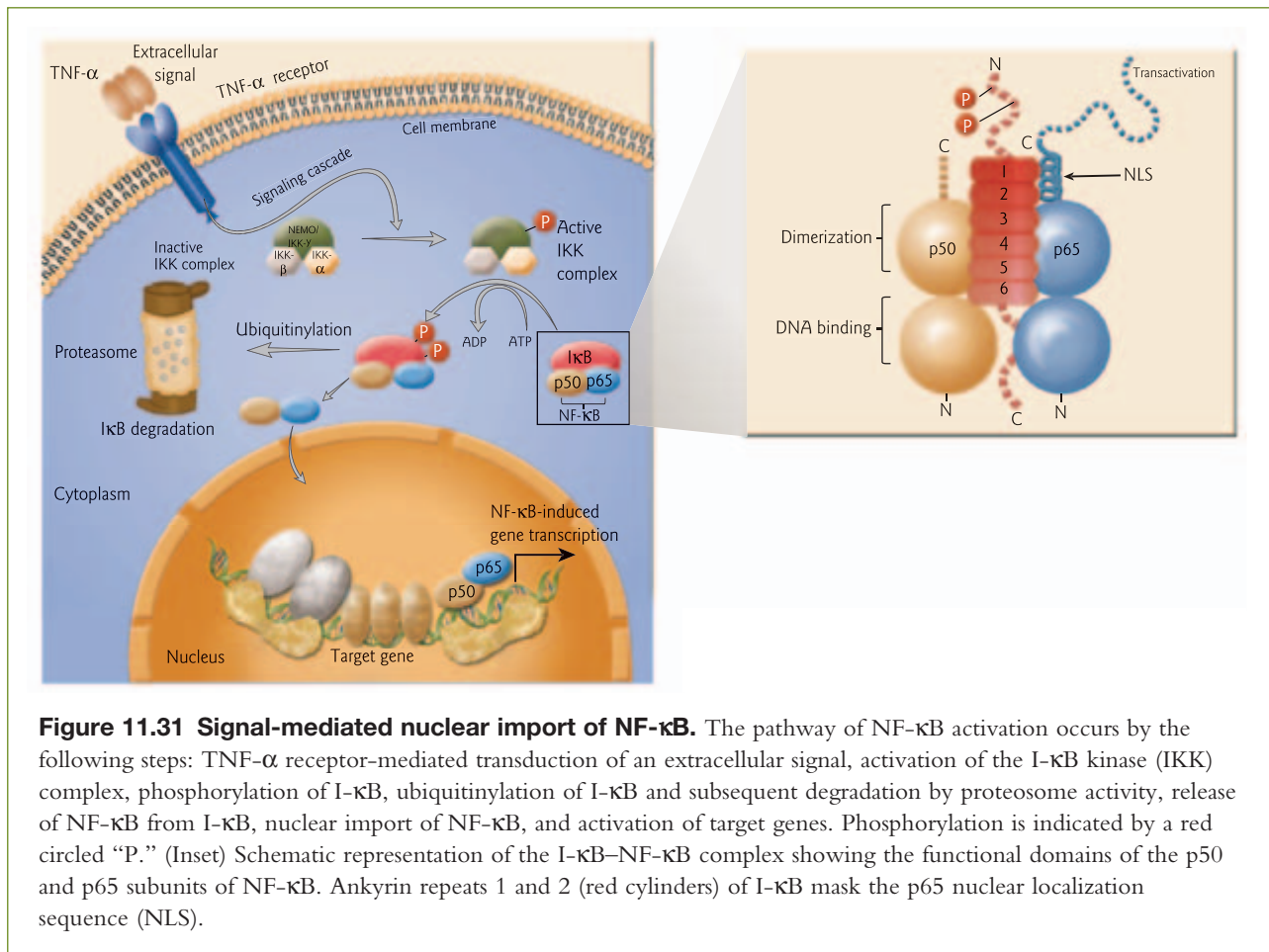
There are various ways that signals are detected by a cell and communicated to a gene. The effect may be direct, where a small molecule, such as a sugar or steroid hormone, enters the cell and binds the transcriptional regulator directly. Or the effect of the signal may be indirect where the signal induces a kinase that phosphorylates the transcriptional regulator or an associated inhibitory protein. This type of indirect signaling is an example of a signal transduction pathway. The signal-mediated nuclear import of transcription factors NF- $\kappa$ B and the glucocorticoid receptor are used to illustrate the wide variety of mechanisms for controlling gene activity.

### Regulated nuclear import of NF- $\kappa$ B

NF- $\kappa$ B (nuclear factor of kappa light polypeptide gene enhancer in B cells) is a dimeric transcription factor that is a central mediator of the human stress response. It plays a key role in regulating cell division, apoptosis, and immune and inflammatory responses. The discovery of NF- $\kappa$ B attracted widespread interest because of the variety of extracellular stimuli that activate it, the diverse genes and biological responses that it controls, and the striking evolutionary conservation of structure and function among family members. NF- $\kappa$ B was first identified in 1986 as a transcription factor in the nuclei of mature B lymphocytes that binds to a 10 bp DNA element in the kappa ( $\kappa$ ) immunoglobulin light-chain enhancer. Since then a family of various distinct subunits has been identified. The events leading to signal-mediated nuclear import of NF- $\kappa$ B involve three main stages: (i) cytoplasmic retention by I- $\kappa$ B; (ii) a signal transduction pathway that induces phosphorylation and degradation of I- $\kappa$ B; and (iii) I- $\kappa$ B degradation resulting in exposure of the NLS on NF- $\kappa$ B, allowing nuclear import of NF- $\kappa$ B (Fig. 11.31).

#### Cytoplasmic retention by I- $\kappa$ B

In a resting B lymphocyte, NF- $\kappa$ B subunits form homodimers or heterodimers in the cytoplasm. The dimers are composed of a DNA-binding subunit (e.g. p50) and a transcription-activating subunit (e.g. p65); p65 also has a DNA-binding domain and binds DNA as a homodimer. In cells that have not received an external cue, the dimers are retained in the cytoplasm in an inactive form. Typically, they are held by an anchor protein called I- $\kappa$ B (inhibitor of  $\kappa$ B) (Fig. 11.31). The I- $\kappa$ B family includes at least eight structurally related proteins, of which I- $\kappa$ B $\alpha$  is the most abundant inhibitor protein. I- $\kappa$ B characteristically contains a



stretch of 5–7 ankyrin repeat domains that mask the NLS within NF-κB. These repeats (named for their discovery in the cytoskeletal protein ankyrin) consist of a pair of antiparallel  $\alpha$ -helices stacked side by side, which are connected by a series of intervening  $\beta$ -hairpin turn motifs.

### Signal transduction pathways induce phosphorylation and degradation of I-κB

Upon receipt of an extracellular signal, a signal transduction pathway is triggered that ultimately leads to transient activation of the serine-specific I-κB kinase (IKK) complex (Fig. 11.31). There are over 150 different possible signals or initiating ligands, including interleukins, bacterial lipopolysaccharides, lymphotoxin- $\beta$ , tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ), viruses, various stress stimuli, and chemotherapeutic agents. Whatever the signal, it is detected by a specific cell surface receptor. The ligand binds to an extracellular domain of the receptor and this binding is communicated to the intracellular domain to initiate a signaling cascade (or second messenger system). There are many complex steps involved in this cascade depending on the particular signal and receptor. For example, the TNF- $\alpha$  signal is relayed through a cascade of kinases ending with phosphorylation and activation of IKK. IKK is composed of two catalytic subunits, IKK $\alpha$  and IKK $\beta$ , and a structural/regulatory subunit IKK $\gamma$  (also called NEMO, not after the clownfish in the Disney movie, but for “NF-κB essential modulator”). The IKK $\beta$  subunit phosphorylates I-κB at two conserved serines (position 32 and 36) in the N-terminus. This post-translational modification leads to release of I-κB from NF-κB.

### I- $\kappa$ B degradation results in exposure of the NLS on NF- $\kappa$ B

Originally, it was postulated that phosphorylation simply caused a conformational change leading to release of I- $\kappa$ B from NF- $\kappa$ B. In fact, phosphorylation of I- $\kappa$ B also triggers another pathway, that of ubiquitinylation and proteasome-mediated degradation (Fig. 11.31; see also Fig. 5.17). Upon release from I- $\kappa$ B, the NLS of NF- $\kappa$ B is exposed. Once the NLS is unmasked, NF- $\kappa$ B can interact with importin- $\alpha/\beta$ 1 and translocate through the NPC. In the nucleus, NF- $\kappa$ B activates target genes by binding to specific DNA regulatory elements. There is an impressively broad range of over 150 target genes with NF- $\kappa$ B-binding sites, including the enhancer of the interferon- $\beta$  gene (see Fig. 11.24) and the kappa ( $\kappa$ ) chain gene involved in the immune response.

### Regulated nuclear import of the glucocorticoid receptor

In the preceding example of NF- $\kappa$ B, gene expression was induced in response to a signal received by a cell surface receptor, and transmitted by a signal transduction cascade. In contrast, the glucocorticoid receptor mediates a highly abbreviated signal transduction pathway: the receptor for the extracellular signal is cytoplasmic and carries the signal directly into the nucleus (Fig. 11.32).

Steroid hormone receptors, such as the glucocorticoid receptor, activate gene expression in response to hormones. Steroid hormones (e.g. cortisol) do not require a cell surface receptor. Because of their lipophilic nature, they can pass through the cell membrane by diffusion. Once in the cytoplasm, they bind the cytoplasmic receptor. In the absence of hormone, the glucocorticoid receptor is bound to heat shock protein 90 (Hsp90) and a 59 kD protein in a complex in the cytoplasm. Hsp90 was first identified as a protein synthesized in response to stress, hence its name. Hsp90 binds to the glucocorticoid receptor via a C-terminal region of the receptor that also binds steroid hormone and masks the region of the glucocorticoid receptor required for dimerization or DNA binding. In the classical model, activation of the glucocorticoid receptor involves ligand-induced conformational changes that result in rapid dissociation of the inhibitory Hsp90 protein. Subsequently, two glucocorticoid receptors join together to form a homodimer. The NLS interacts with importins and the receptor is translocated through the nuclear pore complex into the nucleus. However, some recent studies suggest that continued association with Hsp90 is critical for efficient nuclear import. Thus GR interactions with Hsp90 after ligand binding requires further investigation. Once in the nucleus, the glucocorticoid receptor dimer binds DNA at a glucocorticoid responsive element (GRE) (see Fig. 11.16). Activation of hormone-responsive target genes leads to many diverse cellular responses, ranging from increases in blood sugar to anti-inflammatory action.

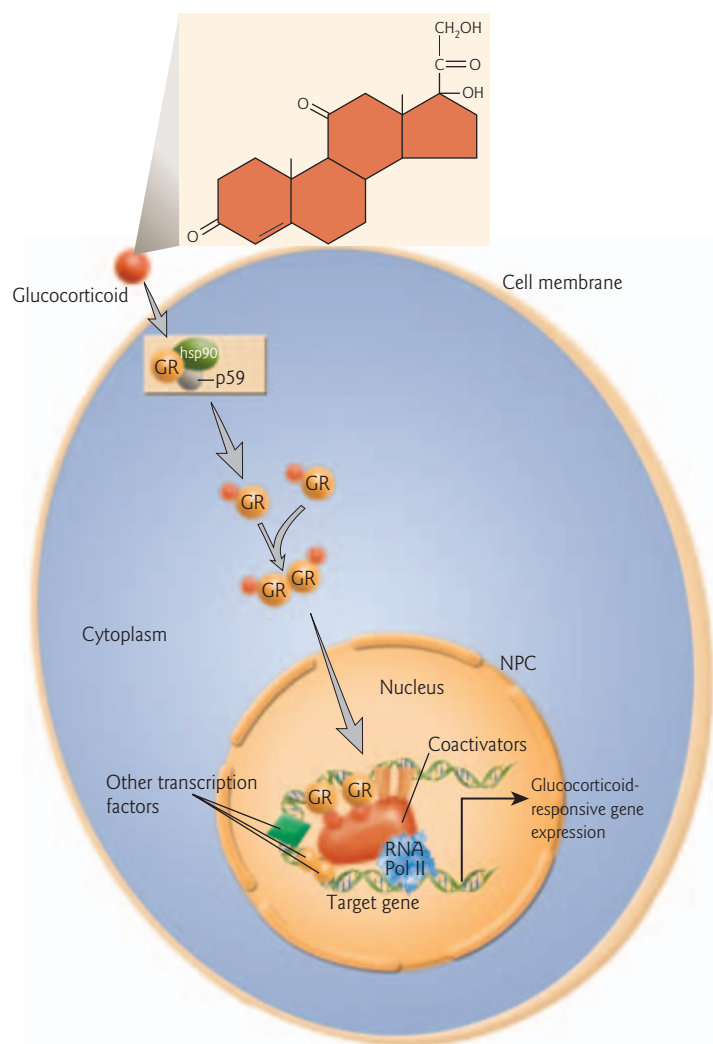
## Chapter summary

The synthesis of tens of thousands of different eukaryotic mRNAs is carried out by RNA polymerase II (RNA pol II). During the process of transcription, RNA pol II associates transiently not only with the template DNA but with many different proteins, including general transcription factors. Transcription factors are sequence-specific DNA-binding proteins that bind to gene promoters and other regulatory elements, interpret the information present in these regulatory elements, and transmit the appropriate response to the RNA pol II transcriptional machinery. The gene promoter consists of core promoter elements and proximal promoter elements that are required for the initiation of transcription or that increase the frequency of initiation only when positioned near the transcriptional start site. The initiation step alone involves the assembly of dozens of factors to form a preinitiation complex. The general transcription factor TFIID is responsible for the recognition of most of the known core promoter elements, the best characterized of which is the TATA box. Long-range regulatory DNA elements act over distances of 100 kb or more from the gene promoter. Enhancers increase gene promoter activity downstream, upstream, or in either orientation relative to the promoter. Similar elements that repress gene activity are called silencers. Insulators act as barriers between regions of heterochromatin and euchromatin and block enhancer or silencer activity of neighboring genes. Locus control regions (LCRs) maintain functional, independent domains of active



**Figure 11.32 Signal-mediated import of the glucocorticoid receptor (GR).** In the

classical model, activation of the GR by glucocorticoid involves its release from the inhibitory Hsp90–p59 complex, allowing dimerization of GR and nuclear import. However, recent studies point towards a role of Hsp90 and other chaperones in efficient nuclear import of GR.



chromatin. They are often characterized by DNase I hypersensitive sites which are indicative of an “open” chromatin conformation. The LCR of the  $\beta$ -globin gene cluster is required for high-level transcription of all genes in the cluster, but chromatin loop formation controls their developmental expression. Matrix attachment regions (MARs) organize chromatin into independent loop domains, and localize transcription factors and actively transcribed genes to the nuclear matrix.

Transcription is mediated by the collective action of sequence-specific DNA-binding transcription factors along with the general RNA pol II transcriptional machinery, an assortment of coregulators that bridge the DNA-binding factors to the transcriptional machinery, a number of chromatin-remodeling factors that mobilize nucleosomes, and a variety of enzymes that catalyze covalent modification of histones and other proteins.

RNA pol II is a 12-subunit polymerase, capable of synthesizing RNA and proofreading nascent transcript. A set of five general transcription factors, denoted TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH, is responsible for promoter recognition and for unwinding the promoter DNA. RNA pol II is absolutely dependent on these auxiliary transcription factors for the initiation of transcription. Assembling the transcription preinitiation complex involves a series of highly ordered steps. Binding of TFIID (TBP and associated TAFs)

to the TATA box provides a platform to recruit other general transcription factors and RNA pol II to the gene promoter. *In vitro*, these proteins assemble at the promoter in the following order: TFIIB (orients the complex on the promoter), TFIIF together with RNA pol II, and then TFIIE and TFIIH. TFIIH has both kinase and helicase activity. Phosphorylation of the C-terminal domain (CTD) of RNA pol II is required for transcript elongation. For recycling of RNA pol II and reinitiation of transcription, the CTD must again be dephosphorylated. Mediator is a 20-subunit complex, which acts as a molecular bridge to transmit regulatory information from activator and repressor proteins to RNA pol II.

Transcription factors mediate gene-specific transcriptional activation or repression. They are modular proteins consisting of a number of domains, including a DNA-binding domain, a transactivation domain that activates transcription via protein–protein interactions, and a dimerization domain. The majority of transcription factors bind DNA as homodimers or heterodimers. Some of the more common DNA-binding domain motifs are the helix–turn–helix, zinc finger, basic leucine zipper, and basic helix–loop–helix motifs. Despite differences in overall three-dimensional structure, all these common motifs involve interaction between an  $\alpha$ -helical domain of the protein and about five base pairs within the major groove of the DNA double helix. The homeodomain is a variant of the helix–turn–helix motif found in many transcription factors that are important in development. The zinc finger is formed by interspersed cysteines and/or histidines that covalently bind a central zinc ion, folding a short length of the amino acid chain into a compact loop domain, one side of which is an  $\alpha$ -helix. The leucine zipper and helix–loop–helix motifs do not directly contact the DNA, instead they play an indirect structural role in DNA binding by facilitating dimerization of two similar transcription factors with basic  $\alpha$ -helical DNA-binding domains.

Transcriptional coactivators and corepressors are proteins that increase or decrease transcriptional activity through protein–protein interactions without binding DNA directly. There are two main classes: (i) chromatin modification complexes that modify histones post-translationally, in ways that allow greater access of other proteins to DNA; and (ii) chromatin remodeling complexes, such as the yeast SWI/SNF family and related families (ISWI and SWR1) that contain ATP-dependent DNA unwinding activities. Chromatin modification complexes include histone acetyltransferases that direct acetylation of lysine residues in histones, and histone methyltransferases that direct methylation of lysines and arginines. These processes are dynamic and are subject to regulation by histone deacetylases and demethylases. Histone modifications provide binding sites for proteins that can change the chromatin state to either active or repressed. The histone code hypothesis proposes that covalent post-translational modifications of histone tails are read by the cell and lead to a complex, combinatorial transcriptional output. Other important histone modifications include ubiquitylation, phosphorylation, ADP-ribosylation, and sumoylation. In some cases, the selection of linker histone variants is important for gene regulation. Chromatin remodeling complexes can mediate at least four different changes in chromatin structure: (i) nucleosome sliding in which the position of a nucleosome on the DNA changes; (ii) remodeled nucleosomes in which the DNA becomes more accessible but the histones remain bound; (iii) complete dissociation of the DNA and histones; or (iv) the replacement of a core histone with a variant histone. Each family of ATP-dependent chromatin remodeling complexes is defined by a unique subunit composition, the presence of a distinct ATPase, and their precise mechanisms for remodeling chromatin.

The particular order of recruitment of the preinitiation complex, the transcription factors, and the chromatin remodeling and modification complexes is gene-specific. The enhanceosome model proposes that interactions among transcription factors promote their cooperative, step-wise assembly on DNA and give the complex exceptional stability. The hit and run model proposes that transcriptional activation reflects the probability that all components required for activation will meet at a certain chromatin site; i.e. transcription complexes are assembled in a stochastic fashion from freely diffusible proteins, and their binding is highly transient and dynamic. Most likely, transcription complex formation is a combination of elements of both models. The principles of combinatorial interaction and complex stability apply even if the complex itself has a very limited lifetime.

After assembly of the preinitiation complex and the rest of the transcriptional machinery, a period of abortive initiation follows before the polymerase escapes the promoter region and enters the elongation

phase. Promoter clearance requires phosphorylation of the CTD of RNA pol II at multiple sites within a series of heptapeptide repeats. Most of the general transcription factors are released. TFIID remains bound to the gene promoter to allow rapid formation of a new preinitiation complex. Once phosphorylated, RNA pol II can unwind DNA without the help of the helicase activity of TFIIH. The elongation phase is subdivided into multiple stages: the selection of a NTP complementary to the DNA template, the catalysis of phosphodiester bond formation, and the translocation of the RNA transcript and DNA template. The growing RNA remains at a single “tunable” active site in RNA pol II that switches between RNA synthesis and cleavage during proofreading of misincorporated nucleotides. The two reactions are mediated by differential coordination of one of the two metal ions held in the active site.

Movement of RNA pol II through the array of nucleosomes on the DNA template occurs by histone H2A-H2B dimer removal. RNA pol III uses a different mechanism which involves nucleosome mobilization – the transfer of the histone core octamer along the DNA without displacement. H2A-H2B dimer removal is mediated by a number of auxiliary elongation factors, including FACT, Elongator, and TFIIS.

Since protein synthesis occurs in the cytoplasm, transcriptional regulatory proteins must be delivered to their site of activity in the nucleus. Trafficking between the nucleus and the cytoplasm occurs via the nuclear pore complexes (NPCs). The NPC is a large multiprotein complex which consists of eight spokes that form a cylinder embedded in the nuclear envelope. The NPCs allow bidirectional passive diffusion of ions and small molecules. Nuclear proteins, RNAs, and RNPs selectively enter and exit the nucleus by a signal-mediated and energy-dependent mechanism. Proteins are targeted to the nucleus by a basic amino acid sequence (typically lysine and arginine rich) called a nuclear localization sequence (NLS). Nuclear proteins that shuttle between the nucleus and cytoplasm also have a nuclear export sequence (NES); the most common of these has a leucine-rich consensus sequence. Nuclear import and export pathways are mediated by a family of soluble receptors referred to as importins or exportins, and collectively called karyopherins. There are multiple pathways for nuclear localization mediated by several different classes of NLSs and NESs and multiple types of karyopherins. Bidirectional nucleocytoplasmic transport through the NPC occurs against a concentration gradient. The energy source and the directional cue are both provided by the small GTPase Ran.

In the nuclear import cycle, an importin binds to its NLS-bearing cargo in the cytoplasm and translocates through the NPC into the nucleus. The exact mechanism by which cargo translocation occurs is poorly understood. In the nucleus, the importin binds RanGTP, resulting in cargo release. The importin-RanGTP complex recycles back to the cytoplasm. After translocation to the cytoplasm, RanGAP mediates GTP hydrolysis on Ran, resulting in dissociation of the complex. The transporter NTF2 carries RanGDP into the nucleus. The conversion of RanGDP to RanGTP is mediated by chromatin-bound RanGEF. In the nuclear export cycle, exportins bind to their cargo in the nucleus in the presence of RanGTP. In the cytoplasm, GTP hydrolysis causes disassembly of the export complex and recycling of the export receptor.

Regulation of transcription factor nuclear import by signal transduction pathways is a powerful level of gene regulation. The events leading to the signal-mediated nuclear import of NF- $\kappa$ B, a transcription factor that is a central mediator of the human stress response, involve three main stages: (i) cytoplasmic retention by the inhibitor protein I- $\kappa$ B; (ii) a signal transduction pathway triggered by an extracellular signal (e.g. bacterial lipopolysaccharides or TNF- $\alpha$ ) that induces phosphorylation and degradation of I- $\kappa$ B; and (iii) I- $\kappa$ B degradation resulting in exposure of the NLS on NF- $\kappa$ B, allowing nuclear import of NF- $\kappa$ B. Once in the nucleus, NF- $\kappa$ B activates target genes by binding to specific DNA regulatory elements.

The glucocorticoid receptor (GR) mediates a highly abbreviated signal transduction pathway leading to many diverse cellular responses, ranging from increases in blood sugar to anti-inflammatory action. In the absence of hormones, GR is bound in a complex with Hsp90 and p59. Hsp90 masks the regions of GR necessary for formation of homodimers and DNA binding. In the classical model, activation of GR by a glucocorticoid involves a ligand-induced conformational change that releases it from the Hsp90-p59 complex, allowing dimerization and nuclear import. Recent evidence suggests, however that Hsp90 may play a role in efficient nuclear entry. Once in the nucleus, the GR activates target genes by binding to specific DNA regulatory elements.

## Analytical questions

- 1 You suspect that a sequence upstream of a transcriptional start site is acting as an enhancer and not as a promoter. Describe an experiment you would run to test your hypothesis. Predict the results.
- 2 A team of molecular biologists has made transgenic mice using a viral promoter and a cDNA coding for a human protein called p45. In most of the transgenic mice, they observe position-dependent expression of the human gene. They seek your advice. Suggest regulatory regions that you could link to the promoter–p45 DNA construct that might confer position-independent expression. Explain your choice of regulatory regions.
- 3 You have purified a transcription factor that has a leucine-rich region. You perform an electrophoretic mobility shift assay (EMSA) using a double-stranded oligonucleotide that you know from other studies contains the site recognized by this transcription factor *in vivo*. However, the transcription factor does not bind to the labeled oligonucleotide in your EMSA. Provide an explanation for this result.
- 4 You are studying a new class of eukaryotic promoters recognized by a novel RNA polymerase. You discover two general transcription factors that are required for transcription of these promoters. You suspect that one has helicase activity and that the other is required to recruit the helicase and the RNA polymerase to the promoter. Describe experiments you would perform to test your hypothesis. Provide sample results of your experiments.
- 5 The figure below shows an electrophoretic mobility shift assay (EMSA). The plus and minus symbols indicate which of the following components were included in each shift assay:

Naked DNA =  $^{32}\text{P}$ -labeled DNA fragment containing a promoter element

Mononucleosome = same  $^{32}\text{P}$ -labeled DNA wrapped around the histone core octamer

ZF = zinc finger protein

$\Delta\text{ZF}$  = ZF with the zinc finger deleted

SWI/SNF = SWI/SNF complex

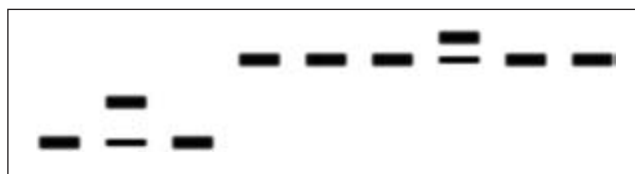
HAT = histone acetyltransferase

HDAC = histone deacetylase

Naked DNA	+	+	+	–	–	–	–	–	–
Mononucleosome	–	–	–	+	+	+	+	+	+
ZF	–	+	–	–	+	+	+	+	+
$\Delta\text{ZF}$	–	–	+	–	–	–	–	–	–
SWI/SNF	–	–	–	–	–	+	+	–	+
HAT	–	–	–	–	–	–	+	+	–
HDAC	–	–	–	–	–	–	–	–	+

Sample wells

End of Gel



Interpret the results for each lane.

- 6 You include an inhibitor of the protein kinase activity of TFIIF in an *in vitro* transcription assay. What step in transcription would you expect to see blocked? Describe an experiment you would run to test your hypothesis. Predict the results.
- 7 Assume that you are studying a transcription factor (TF). You have constructed a plasmid vector for the expression of a GFP-tagged TF. You discover that the TF is sometimes in the nucleus and sometimes in the cytoplasm. Describe a possible mechanism for its cytoplasmic retention and subsequent signal-mediated import. Design an experiment to test your hypothesis and provide sample results.

## Suggestions for further reading

- Adelman, K., Lis, J.T. (2002) How does pol II overcome the nucleosome barrier? *Molecular Cell* 9:451–452.
- Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., Thanos, D. (2000) Ordered recruitment of chromatin modifying and general transcription factors to the IFN- $\beta$  promoter. *Cell* 103:667–678.
- Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V.R., Bianchi, M.E. (2005) GR and HMGB1 interact only within chromatin and influence each other's residence time. *Molecular Cell* 18:109–121.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiology Reviews* 29:231–262.
- Armache, K.J., Kettenberger, H., Cramer, P. (2005) The dynamic machinery of mRNA elongation. *Current Opinion in Structural Biology* 15:197–203.
- Baumli, S., Hoepfner, S., Cramer, P. (2005) A conserved Mediator hinge revealed in the structure of the MED7/MED21 (Med7/Srb7) heterodimer. *Journal of Biological Chemistry* 280:18171–18178.
- Bell, A.C., West, A.G., Felsenfeld, G. (2001) Insulators and boundaries. Versatile regulatory elements in the eukaryotic genome. *Science* 291:447–450.
- Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., Reinberg, D. (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science* 301:1090–1093.
- Bird, R.C., Stein, G.S., Lian, J.B., Stein, J.L. (1997) *Nuclear Structure and Gene Expression*. Academic Press, New York.
- Bode, J., Goetze, S., Heng, H., Krawetz, S.A., Benham, C. (2003) From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Research* 11:435–445.
- Boeger, H., Bushnell, D.A., Davis, R. et al. (2005) Structural basis of eukaryotic gene transcription. *FEBS Letters* 579:899–903.
- Boeger, H., Griesenbeck, J., Strattan, J.S., Kornberg, R.D. (2004) Removal of promoter nucleosomes by disassembly rather than sliding *in vivo*. *Molecular Cell* 14:667–673.
- Bolzer, A., Kreth, G., Solovei, I. et al. (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology* 3:826–842.
- Brand, M., Ranish, J.A., Kummer, N.T. et al. (1994) Sequences containing the second-intron enhancer are essential for transcription of the human apolipoprotein B gene in the livers of transgenic mice. *Molecular and Cellular Biology* 14:2243–2256.
- Calladine, C.R., Drew, H.R., Luisi, B.F., Travers, A.A. (2004) *Understanding DNA. The Molecule and How it Works*, 3rd edn. Elsevier Academic Press, New York.
- Carroll, S.B. (2005) *From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design*, 2nd edn. Blackwell Publishing, Malden, MA.
- Chubb, J.R., Bickmore, W.A. (2003) Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* 112:403–406.
- Cirillo, L., Zaret, K. (2004) Developmental biology. A linker histone restricts muscle development. *Science* 304:1607–1609.
- Cook, P.R. (2001) *Principles of Nuclear Structure and Function*. Wiley-Liss, New York.
- Corda, D., Di Girolamo, M. (2003) Functional aspects of protein mono-ADP-ribosylation. *EMBO Journal* 22:1953–1958.
- Crabtree, G.R., Aebersold, R., Groudine, M. (2004) Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nature Structural Molecular Biology* 11:73–80.
- Cramer, P. (2004) RNA polymerase II structure: from core to functional complexes. *Current Opinion in Genetics and Development* 14:218–226.



- Cuajungco, M.P., Leyne, M., Mull, J. et al. (2003) Tissue-specific reduction in splicing efficiency of *IKBKAP* due to the major mutation associated with familial dysautonomia. *American Journal of Human Genetics* 72:749–758.
- de la Cruz, X., Lois, S., Sánchez-Molina, S., Martínez-Balbás, M.A. (2005) Do protein motifs read the histone code? *BioEssays* 27:164–175.
- De Sandre-Giovannoli, A., Bernard, R., Cau, P. et al. (2003) Lamin A truncation in Hutchinson–Gilford progeria. *Science* 300:2055.
- Dion, M.F., Altschuler, S.J., Wu, L.F., Rando, O.J. (2005) Genomic characterization reveals a simple histone H4 acetylation code. *Proceedings of the National Academy of Sciences USA* 102:5501–5506.
- Flanagan, P.M., Kelleher, R.J. III, Sayre, M.H., Tschochner, H., Kornberg, R.D. (1991) A mediator required for activation of RNA polymerase II transcription *in vitro*. *Nature* 350:436–438.
- Fry, C.J., Peterson, C.L. (2002). Unlocking the gates to gene expression. *Science* 295:1847–1848.
- Green, M.R. (2005) Eukaryotic transcription activation: right on target. *Molecular Cell* 18:399–402.
- Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T., Adam, S.A. (2004) Importin  $\alpha$ : a multipurpose nuclear-transport receptor. *Trends in Cell Biology* 14:505–514.
- Guzder, S.N., Sung, P., Bailly, V., Prakash, L., Prakash, S. (1994) RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369:578–581.
- Hay, R.T. (2005) SUMO: a history of modification. *Molecular Cell* 18:1–12.
- Heng, H.H.Q., Goetze, S., Ye, C.J. et al. (2004) Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *Journal of Cell Science* 117:999–1008.
- Kadonaga, J.T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA-binding factors. *Cell* 116:247–257.
- Kalderon, D., Richardson, W.D., Markham, A.F., Smith, A.E. (1984) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* 311:33–38.
- Kassabov, S.R., Zhang, B., Persinger, J., Bartholomew, B. (2003) SWI/SNF unwraps, slides, and rewraps the nucleosome. *Molecular Cell* 11:391–403.
- Kettenberger, H., Armache, K.J., Cramer, P. (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular Cell* 16:955–965.
- Kim, J.H., Lane, W.S., Reinberg, D. (2002) Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proceedings of the National Academy of Sciences USA* 99:1241–1246.
- Korber, P., Hörz, W. (2004) SWRred not shaken: mixing the histones. *Cell* 117:5–7.
- Kouzarides, T., Ziff, E. (1988) The role of the leucine zipper in the fos–jun interaction. *Nature* 336:646–651.
- Kurdistani, S.K., Tavazoie, S., Grunstein, M. (2004) Mapping global histone acetylation patterns to gene expression. *Cell* 117:721–733.
- Kutay, U., Güttinger, S. (2005) Leucine-rich nuclear export signals: born to be weak. *Trends in Cell Biology* 15:121–124.
- Latchman, D. (2004) *Eukaryotic Transcription Factors*, 4th edn. Elsevier Academic Press, New York.
- Laudet, V., Gronemeyer, H. (2002) *The Nuclear Receptor Facts Book*. Academic Press, New York.
- Levine, M., Tjian, R. (2003) Transcription regulation and animal diversity. *Nature* 424:147–151.
- Lewis, B.A., Sims, R.J. III, Lane, W.S., Reinberg, D. (2005) Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. *Molecular Cell* 18:471–481.
- Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., Kadonaga, J.T. (2004) The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes and Development* 18:1606–1617.

- Low, L.Y., Hernández, H., Robinson, C.V., O'Brien, R., Grossmann, J.G., Ladbury, J.E., Luisi, B. (2002) Metal-dependent folding and stability of nuclear hormone receptor DNA-binding domains. *Journal of Molecular Biology* 319:87–106.
- Lu, H., Zawel, L., Fisher, L., Egly, J.M., Reinberg, D. (1992) Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358:641–645.
- MacLean, J.A. II, Chen, M.A., Wayne, C.M. et al. (2005) *Rhox*: a new homeobox gene cluster. *Cell* 120:369–382.
- Marx, J. (2005) Combing over the Polycomb group proteins. *Science* 308:624–626.
- Matsuura, Y., Stewart, M. (2004) Structural basis for the assembly of a nuclear export complex. *Nature* 432:872–877.
- McKnight, R.A., Shamay, A., Sankaran, L., Wall, R.J., Hennighausen, L. (1992) Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proceedings of the National Academy of Sciences USA* 89:6943–6947.
- McKnight, S.L. (1991) Molecular zippers in gene regulation. *Scientific American* 265:32–39.
- Mohrmann, L., Verrijzer, C.P. (2005) Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochimica et Biophysica Acta* 1681:59–73.
- Mosammaparast, N., Pemberton, L.F. (2004) Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends in Cell Biology* 14:547–556.
- Müller, F., Tora, L. (2004) The multicolored world of promoter recognition complexes. *EMBO Journal* 23:2–8.
- Navas, P.A., Li, Q., Peterson, K.R., Swank, R.A., Rohde, A., Roy, J., Stamatoyannopoulos, G. (2002) Activation of the  $\beta$ -like globin genes in transgenic mice is dependent on the presence of the  $\beta$ -locus control region. *Human Molecular Genetics* 11:893–903.
- Owen-Hughes, T., Bruno, M. (2004) Breaking the silence. *Science* 303:324–325.
- Pal, M., Ponticelli, A.S., Luse, D.S. 2005. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Molecular Cell* 19:101–110.
- Pederson, T. (2000) Half a century of “the nuclear matrix.” *Molecular Biology of the Cell* 11:799–805.
- Pemberton, L.F., Paschal, B.M. (2005) Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* 6:187–198.
- Rachez, C., Freedman, L.P. (2001) Mediator complexes and transcription. *Current Opinion in Cell Biology* 13:274–280.
- Rahl, P.B., Chen, C.Z., Collins, R.N. (2005) Elp1p, the yeast homolog of the FD disease syndrome protein, negatively regulates exocytosis independently of transcriptional elongation. *Molecular Cell* 17:841–853.
- Renault, L., Kuhlmann, J., Henkel, A., Wittinghofer, A. (2001) Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). *Cell* 105:245–255.
- Rhoades, D., Klug, A. (1993) Zinc fingers. *Scientific American* 268:32–39.
- Riddick, G., Macara, I.G. (2005) A systems analysis of importin- $\alpha$ - $\beta$  mediated nuclear protein import. *Journal of Cell Biology* 168:1027–1038.
- Sassone-Corsi, P., Corden, J., Kédinger, C., Chambon, P. (1981) Promotion of specific *in vitro* transcription by excised “TATA” box sequences inserted in a foreign nucleotide environment. *Nucleic Acids Research* 9:3941–3958.
- Schwartz, T.U. (2005) Modularity within the architecture of the nuclear pore complex. *Current Opinion in Structural Biology* 15:221–226.
- Semenza, G.L. (1999) *Transcription Factors and Human Disease*. Oxford University Press, New York.
- Shank, L.C., Paschal, B.M. (2005) Nuclear transport of steroid hormone receptors. *Critical Reviews in Eukaryotic Gene Expression* 15:49–73.
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119:941–953.

- Slaugenhaupt, S.A., Gusella, J.F. (2002) Familial dysautonomia. *Current Opinion in Genetics and Development* 12:307–311.
- Smale, S.T., Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annual Review of Biochemistry* 72:449–479.
- Smith, C.L., Horowitz-Scherer, R., Flanagan, J.F., Woodcock, C.L., Peterson, C.L. (2003) Structural analysis of the yeast SWI/SNF chromatin remodeling complex. *Nature Structural Biology* 10:141–145.
- Spiegelman, B.M., Heinrich, R. (2004) Biological control through regulated transcriptional coactivators. *Cell* 119:157–167.
- Spitz, F., Duboule, D. (2005) Developmental biology: reproduction in clusters. *Nature* 434:715–716.
- Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., Felsenfeld, G. (2004) Chromatin remodeling by RNA polymerases. *Trends in Biochemical Sciences* 29:127–135.
- Svejstrup, J.Q. (2003) Histones face the FACT. *Science* 301:1053–1055.
- Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active  $\beta$ -globin locus. *Molecular Cell* 10:1453–1465.
- Vakoc, C.R., Letting, D.L., Gheldof, N. et al. (2005) Proximity among distant regulatory elements at the  $\beta$ -globin locus requires GATA-1 and FOG-1. *Molecular Cell* 17:453–462.
- Vieira, K.F., Levings, P.P., Hill, M.A., Cruselle, V.J., Kang, S.H.L., Engel, J.D., Bungert, J. (2004) Recruitment of transcription complexes to the  $\beta$ -globin gene locus *in vivo* and *in vitro*. *Journal of Biological Chemistry* 279:50350–50357.
- Villavicencio, E.H., Walterhouse, D.O., Iannaccone, P.M. (2000) The Sonic hedgehog-Patched-Gli pathway in human development and disease. *American Journal of Human Genetics* 67:1047–1054.
- Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., Berk, A.J. (2005) Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Molecular Cell* 17:683–694.
- Weeber, E.J., Levenson, J.M., Sweatt, J.D. (2002) Molecular genetics of human cognition. *Molecular Interventions* 2:376–391.
- Weis, K. (2003) Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* 112:441–451.
- Wilson, E.B. (1925) *The Cell in Development and Heredity*. Macmillan Co., New York.
- Wolffe, A. (1998) *Chromatin Structure and Function*, 3rd edn. Academic Press, New York.
- Yie, J., Merika, M., Munshi, N., Chen, G., Thanos, D. (1999) The role of HMG I(Y) in the assembly and function of the INF- $\beta$  enhanceosome. *EMBO Journal* 18:3074–3089.
- Yusufzai, T.M., Tagami, H., Nakatani, Y., Felsenfeld, G. (2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Molecular Cell* 13:291–298.
- Zhou, J., Berger, S.L. (2004) Good fences make good neighbors: barrier elements and genomic regulation. *Molecular Cell* 16:500–502.