Developmental Genetics

2

But whatever the immediate operations of the genes turn out to be, they most certainly belong to the category of developmental processes and thus belong to the province of embryology.

C. H. WADDINGTON (1956)

We have entered the cell, the mansion of our birth, and have started the inventory of our acquired wealth.

ALBERT CLAUDE (1974)

CYTOLOGICAL STUDIES DONE AT THE TURN OF THE TWENTIETH CENTURY established that the chromosomes in each cell of an organism's body are the mitotic descendants of the chromosomes established at fertilization (Wilson 1896; Boveri 1904). In other words, each somatic cell nucleus has the same chromosomes—and therefore the same set of genes—as all the other somatic nuclei. This fundamental concept is called **genomic equivalence**. Given this concept, one of the major questions facing biologists of the early twentieth century was how nuclear genes could direct development when these genes are the same in every cell type (Harrison 1937; Just 1939). If every cell in the body contains the genes for hemoglobin and insulin proteins, why is it that hemoglobin proteins are made only in red blood cells, insulin proteins are made only in certain pancreas cells, and neither protein is made in the kidneys or nervous system?

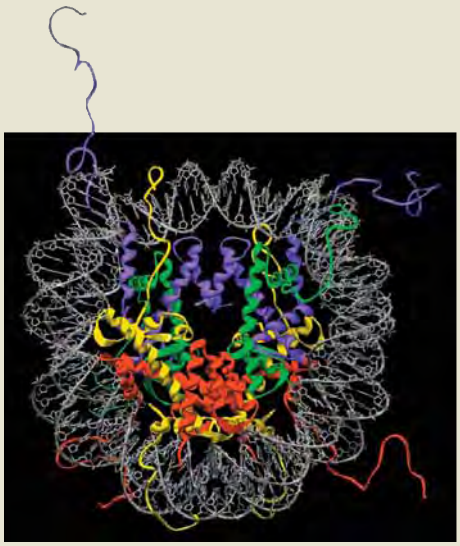
Based on the embryological evidence for genomic equivalence (as well as on bacterial models of gene regulation), a consensus emerged in the 1960s that the answer to this question lies in **differential gene expression**. The three postulates of differential gene expression are:

- Every cell nucleus contains the complete genome established in the fertilized egg. In molecular terms, the DNAs of all differentiated cells are identical.
- The unused genes in differentiated cells are neither destroyed nor mutated, but retain the potential for being expressed.
- Only a small percentage of the genome is expressed in each cell, and a portion of the RNA synthesized in each cell is specific for that cell type.

Gene expression can be regulated at several levels such that different cell types synthesize different sets of proteins:

- **Differential gene transcription** regulates which of the nuclear genes are transcribed into nuclear RNA.
- **Selective nuclear RNA processing** regulates which of the transcribed RNAs (or which parts of such a nuclear RNA) are able to enter into the cytoplasm and become messenger RNAs.
- **Selective messenger RNA translation** regulates which of the mRNAs in the cytoplasm are translated into proteins.
- **Differential protein modification** regulates which proteins are allowed to remain and/or function in the cell.

Some genes (such as those coding for the globin proteins of hemoglobin) are regulated at all these levels.



UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

Evidence for Genomic Equivalence

Until the mid-twentieth century, genomic equivalence was not so much proved as it was assumed (because every cell is the mitotic descendant of the fertilized egg. One of the first tasks of developmental genetics was to determine whether every cell of an organism indeed does have the same *genome*—that is, the same set of genes—as every other cell.

Evidence that every cell in the body has the same genome originally came from the analysis of *Drosophila* chromosomes, in which the DNA of certain larval tissues undergoes numerous rounds of DNA replication without separation, such that the structure of the chromosomes can be seen. In these **polytene** (Gr. "many strands") **chromosomes**, no structural differences were seen between cells; but different regions were seen to be "puffed up" at different times and in different cell types, suggesting that these areas were actively making RNA (Beerman 1952).

See WEBSITE 2.1
Does the genome or the cytoplasm direct development?

See WEBSITE 2.2
The origins of developmental genetics

When Giemsa dyes allowed such observations to be made in mammalian chromosomes, it was also found that no chromosomal regions were lost in most cells. These observations, in turn, were confirmed by nucleic acid hybridization studies, which (for instance) found globin genes in pancreatic tissue, which does not make globin proteins.

But the ultimate test of whether the nucleus of a differentiated cell has undergone irreversible functional restriction is to have that nucleus generate every other type of differentiated cell in the body. If each cell's nucleus is identical to the zygote nucleus, then each cell's nucleus should also be capable of directing the entire development of the organism when transplanted into an activated enucleated egg. As early as 1895, the embryologist Yves Delage predicted that "If, without deterioration, the egg nucleus could be replaced by the nucleus of an ordinary embryonic cell, we should probably see this egg developing without changes" (Delage 1895, p. 738).

In 1952, Briggs and King demonstrated that blastula cell nuclei could direct the development of complete tadpoles when transferred into the cytoplasm of an activated enucleated frog egg. This procedure is called **somatic nuclear transfer** or, more commonly, **cloning**. Nuclei from adult frogs, however, were not able to generate adult frogs. However, adult nuclei (from skin cells, for instance) were

SIDELIGHTS & SPECULATIONS

The Basic Tools of Developmental Genetics

DNA analysis

Embryologist Theodor Boveri (1904) wrote that to discover the mechanisms of development, it was "not cell nuclei, not even individual chromosomes, but certain parts of certain chromosomes from certain cells that must be isolated and collected in enormous quantities for analysis." This analysis was finally made possible by the techniques of gene cloning, DNA sequencing, Southern blotting, gene knockouts, and enhancer traps. In addition, techniques for showing which enhancers and promoters are methylated and which are unmethylated have become more important, as investigations of differential gene transcription have focused on these elements.

For discussions on these techniques, please see the Developmental Biology Website 2.3.

RNA analysis

Differential gene transcription is critical in development. In order to know the time of gene expression and the place of gene expression, one has to be able to have procedures that actually locate a particular type of messenger RNA. These techniques include northern blots, RT-PCR, in situ hybridization, and array technology. To ascertain the function of these mRNAs, new techniques have been formulated, which include antisense and RNA interference (which destroy messages), Cre-lox analysis (which allows the message to be made or destroyed in particular cell types) and ChIP-on-Chip techniques (which enable one to localize active chromatin).

For discussions of these techniques, please see the Developmental Biology Website 2.4.

Bioinformatics

Modern developmental genetics often involves comparing DNA sequences (especially regulatory units such as enhancers and 3′ UTRs) and looking at specific genomes to determine how genes are being regulated. "High-throughput" RNA analysis by micro- and macroarrays enables researchers to compare thousands of mRNAs, and computer-aided synthetic techniques can predict interactions between proteins and mRNAs. Various free websites enable researchers to use the tools that allow such comparisons. Other sites are organism- or organ-specific and are used by researchers studying that particular organ or organism.

For more on these sites and links to them, please see the Developmental Biology Website 2.5

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

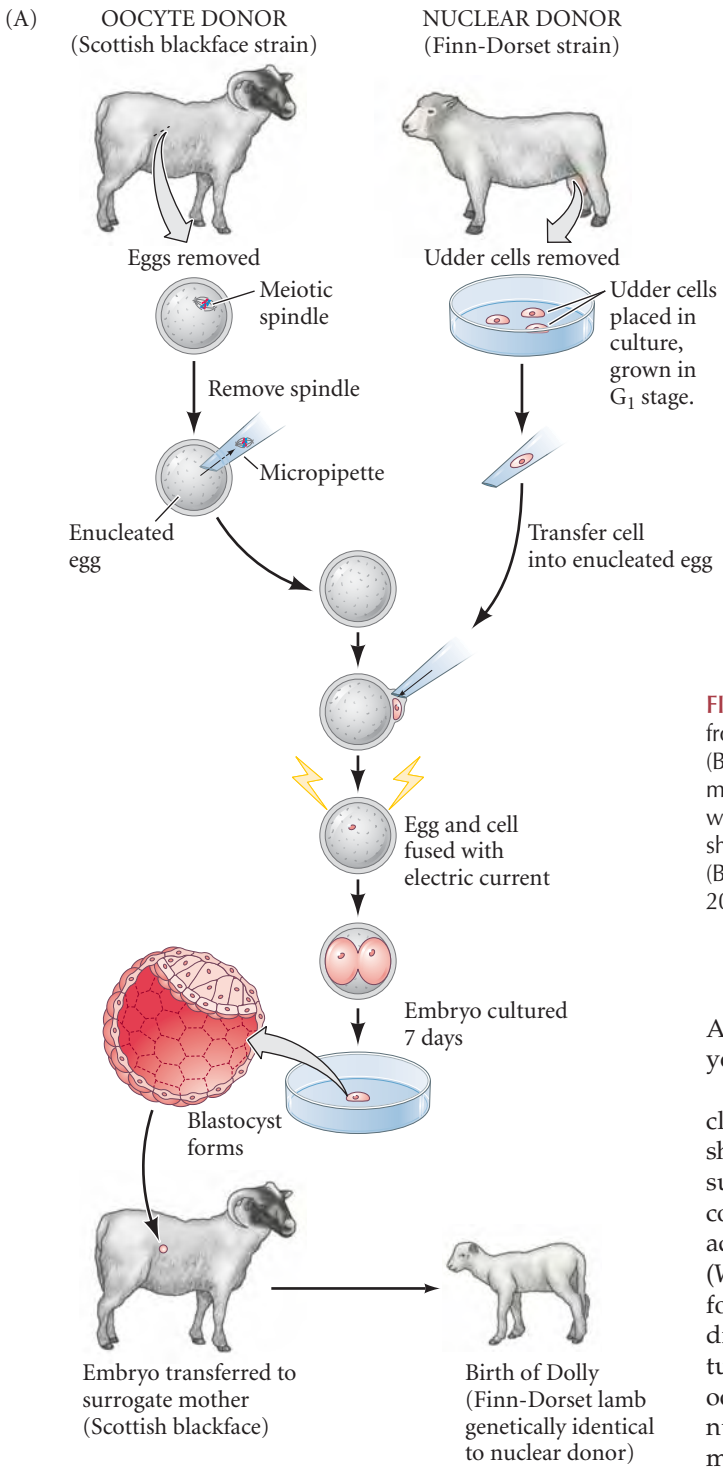


FIGURE 2.1 Cloned mammals have been created using nuclei from adult somatic cells. (A) Procedure used for cloning sheep. (B) Dolly, the adult sheep on the left, was derived by fusing a mammary gland cell nucleus with an enucleated oocyte, which was then implanted in a surrogate mother (of a different breed of sheep) that gave birth to Dolly. Dolly later gave birth to a lamb (Bonnie, at right) by normal reproduction. (A after Wilmut et al. 2000; B, photograph by Roddy Field, © Roslin Institute.)

able to direct the development of all the organs of the tadpoles (Gurdon et al. 1975). Although the tadpoles all died prior to feeding, their existence showed that a single differentiated cell nucleus still retained incredible potencies.

A nucleus of a skin cell could produce all the cells of a young tadpole.

In 1997, Ian Wilmut announced that a sheep had been cloned from a somatic cell nucleus from an adult female sheep. This was the first time an adult vertebrate had been successfully cloned from another adult. Wilmut and his colleagues had taken cells from the mammary gland of an adult (6-year-old) pregnant ewe and put them into culture (Wilmut et al. 1997; Figure 2.1A). The culture medium was formulated to keep the nuclei in these cells at the intact diploid stage (G1) of the cell cycle. This cell-cycle stage turned out to be critical. The researchers then obtained oocytes from a different strain of sheep and removed their nuclei. These oocytes had to be in the second meiotic metaphase (the stage at which they are usually fertilized). Fusion of the donor cell and the enucleated oocyte was accomplished by bringing the two cells together and sending electric pulses through them, destabilizing the cell membranes and allowing the cells to fuse. The same electric pulses that fused the cells activated the egg to begin development. The resulting embryos were eventually transferred into the uteri of pregnant sheep.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

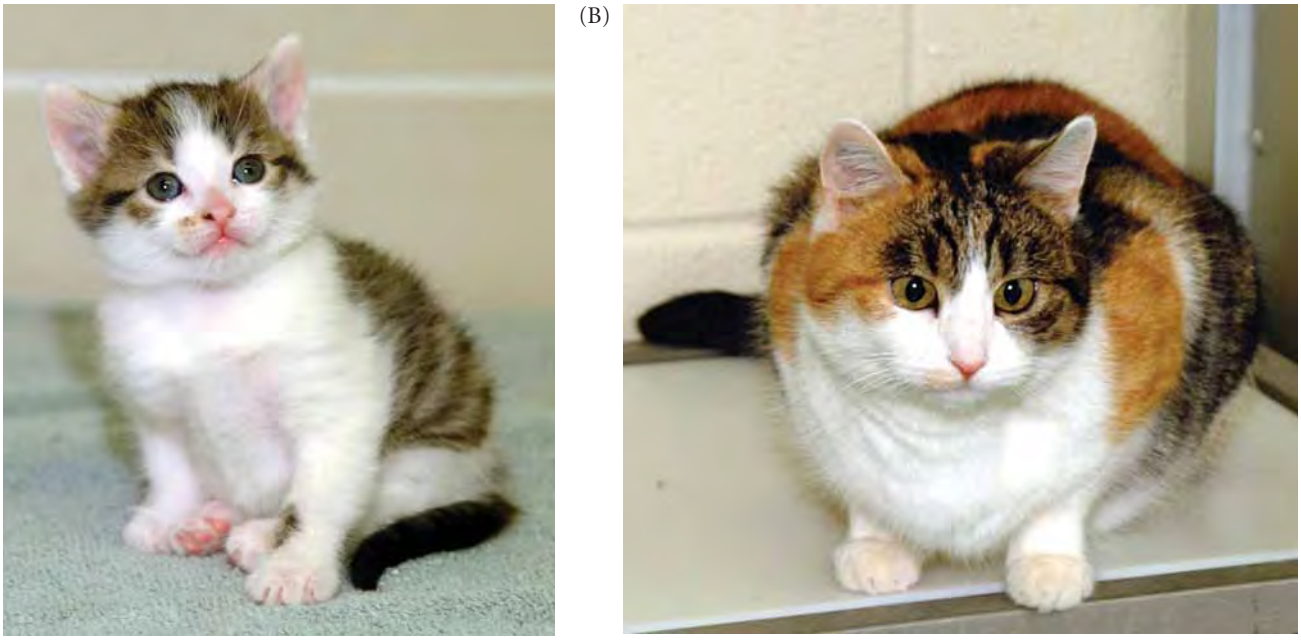


FIGURE 2.2 The kitten "CC" (A) was a clone produced using somatic nuclear transfer from "Rainbow," a female calico cat (B). The two do not appear identical because coat pigmentation pattern in calico cats is affected by the random inactivation of one X chromosome in each somatic cell (see Sidelights & Speculations, p. 50). Their behaviors were also quite different. (Photographs courtesy of College of Veterinary Medicine, Texas A&M University.)

Of the 434 sheep oocytes originally used in this experiment, only one survived: Dolly* (Figure 2.1B). DNA analysis confirmed that the nuclei of Dolly's cells were derived from the strain of sheep from which the donor nucleus was taken (Ashworth et al. 1998; Signer et al. 1998). Cloning of adult mammals has been confirmed in guinea pigs, rabbits, rats, mice, dogs, cats, horses, and cows. In 2003, a cloned mule became the first sterile animal to be so reproduced (Woods et al. 2003). Thus it appears that the nuclei of vertebrate adult somatic cells contain all the genes needed to generate an adult organism. No genes necessary for development have been lost or mutated in the somatic cells.†

Certain caveats must be applied, however. First, although it appears that all the organs were properly formed in the cloned animals, many of the clones developed debilitating diseases as they matured (Humphreys et al. 2001; Jaenisch and Wilmut 2001; Kolata 2001). As we will shortly see, this problem is due in large part to the differences in methylation between the chromatin of the zygote and the differentiated cell. Second, the phenotype of the cloned animal is sometimes not identical to that of the animal from which the nucleus was derived. There is variability due to random chromosomal events and the effects of environment. The pigmentation of calico cats, for instance, is due to the *random inactivation* of one or the other X chromosome (a genetic mechanism that will be discussed later in this chapter) in each somatic cell of the female cat embryo. Therefore, the coat color pattern of the first cloned cat, a calico named "CC," were different from those of "Rainbow," the adult calico whose cells provided the implanted nucleus that generated "CC" (Figure 2.2).

The same genotype gives rise to multiple phenotypes in cloned sheep as well. Wilmut noted that four sheep cloned from blastocyst nuclei from the same embryo "are genetically identical to each other and yet are very different in size and temperament, showing emphatically that an animal's genes do not 'determine' every detail of its physique and personality" (Wilmut et al. 2000, p. 5). Wilmut concludes that for this and other reasons, the "resurrection" of lost loved ones by cloning is not feasible.

SEE WEBSITE 2.6 Cloning and nuclear equivalence

*The creation of Dolly was the result of a combination of scientific and social circumstances. These circumstances involved job security, people with different areas of expertise meeting each other, children's school holidays, international politics, and who sits near whom in a pub. The complex interconnections giving rise to Dolly are told in *The Second Creation* (Wilmut et al. 2000), a book that should be read by anyone who wants to know how contemporary science actually works. As Wilmut acknowledged (p. 36), "The story may seem a bit messy, but that's because life is messy, and science is a slice of life."

†Although cloning humans does not seem feasible at present, each cell of the human body (with just a few exceptions, such as lymphocytes) does appear to contain the same genome as every other cell. As we will see in Chapter 17, adding certain activated transcription factors to ordinary skin fibroblasts will convert them into embryonic stem cells that are indeed capable of generating entire embryos, at least in mice.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

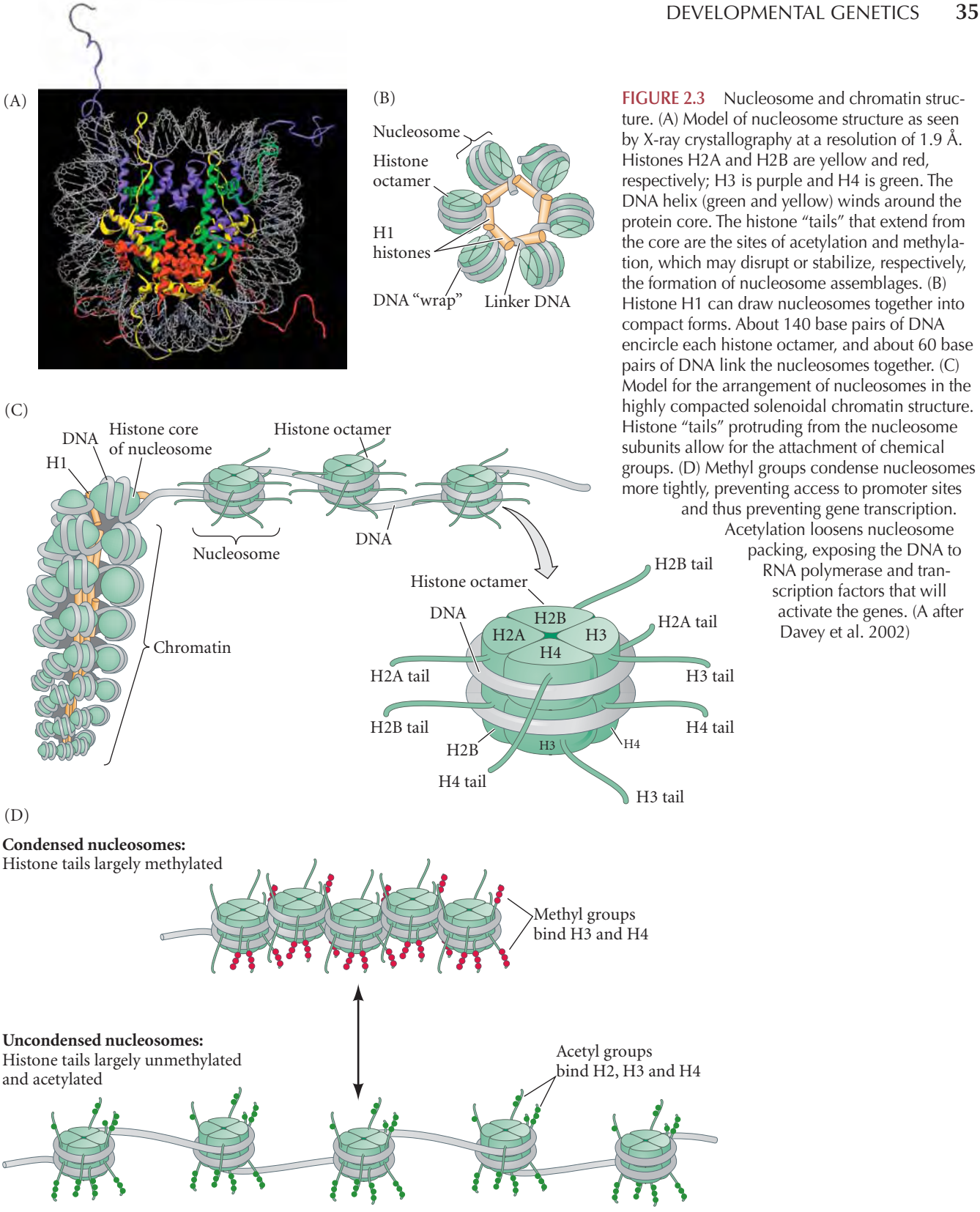


FIGURE 2.3 Nucleosome and chromatin structure. (A) Model of nucleosome structure as seen by X-ray crystallography at a resolution of 1.9 Å. Histones H2A and H2B are yellow and red, respectively; H3 is purple and H4 is green. The DNA helix (green and yellow) winds around the protein core. The histone "tails" that extend from the core are the sites of acetylation and methylation, which may disrupt or stabilize, respectively, the formation of nucleosome assemblages. (B) Histone H1 can draw nucleosomes together into compact forms. About 140 base pairs of DNA encircle each histone octamer, and about 60 base pairs of DNA link the nucleosomes together. (C) Model for the arrangement of nucleosomes in the highly compacted solenoidal chromatin structure. Histone "tails" protruding from the nucleosome subunits allow for the attachment of chemical groups. (D) Methyl groups condense nucleosomes more tightly, preventing access to promoter sites and thus preventing gene transcription. Acetylation loosens nucleosome packing, exposing the DNA to RNA polymerase and transcription factors that will activate the genes. (A after Davey et al. 2002)

Differential Gene Transcription

So how does the same genome give rise to different cell types? To understand this, one needs to understand the anatomy of the genes. One of the fundamental differences distinguishing most eukaryotic genes from prokaryotic genes is that eukaryotic genes are contained within a complex of DNA and protein called **chromatin**. The protein component constitutes about half the weight of chromatin and is composed largely of **histones**. The **nucleosome** is the basic unit of chromatin structure (Figure 2.3). It is composed of an octamer of histone proteins (two molecules

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

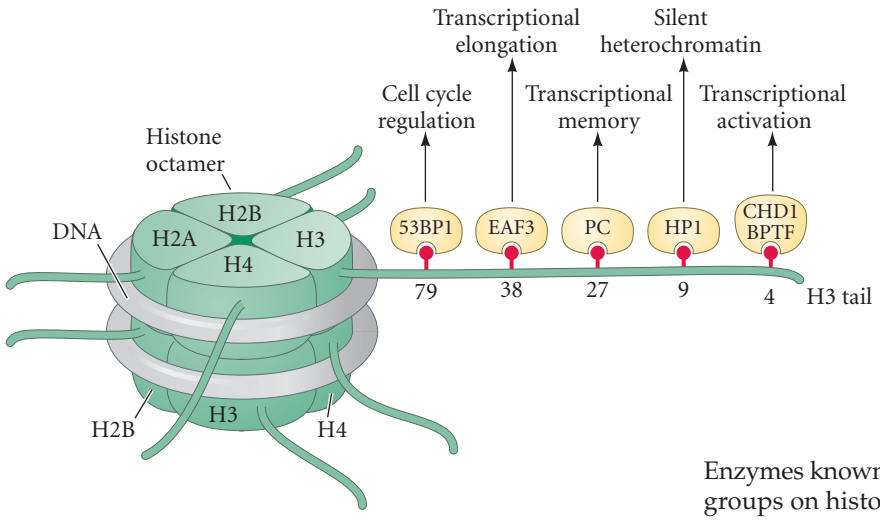


FIGURE 2.4 Histone methylations on histone H3. The tail of histone H3 (its aminomost sequence, at the beginning of the protein) sticks out from the nucleosome and is capable of being methylated or acetylated. Here, lysines can be methylated and recognized by particular proteins. Methylated lysine residues at positions 4, 38, and 79 are associated with gene activation, whereas methylated lysines at positions 9 and 27 are associated with repression. The proteins binding these sites (not shown to scale) are represented above the methyl group. (After Kouzarides and Berger 2007.)

each of histones H2A, H2B, H3, and H4) wrapped with two loops containing approximately 140 base pairs of DNA (Kornberg and Thomas 1974). Histone H1 is bound to the 60 or so base pairs of "linker" DNA between the nucleosomes (Weintraub 1984). There are 14 points of contact between the DNA and the histones (Luger et al. 1997).

Anatomy of the gene: Active and repressed chromatin

Whereas classical geneticists have likened genes to "beads on a string," molecular geneticists liken genes to "string on the beads," an image in which the beads are nucleosomes. Most of the time, the nucleosomes are wound into tight "solenoids" that are stabilized by histone H1 (Figure 2.3C). This H1-dependent conformation of nucleosomes inhibits the transcription of genes in somatic cells by packing adjacent nucleosomes together into tight arrays that prevent transcription factors and RNA polymerases from gaining access to the genes (Thoma et al. 1979; Schlissel and Brown 1984). It is generally thought, then, that the "default" condition of chromatin is a repressed state, and that tissue-specific genes become activated by local interruption of this repression (Weintraub 1985).

HISTONES AS AN ACTIVATION SWITCH The histones are critical because they are responsible for maintaining the repression of gene expression. This repression can be locally strengthened (so that it becomes very difficult to transcribe those genes in the nucleosomes) or relieved (so that transcribing them becomes relatively easy) by modifying the histones (Figure 2.3D). Repression and activation are controlled to a large extent by modifying the tails of histones H3 and H4 with two small organic groups: methyl (CH_3) and acetyl ($COCH_3$) residues. In general, **histone acetylation**—the addition of negatively charged acetyl groups to histones—neutralizes the basic charge of lysine and loosens the histones. This activates transcription. Enzymes known as **histone acetyltransferases** place acetyl groups on histones (especially on lysines in H3 and H4), destabilizing the nucleosomes so that they come apart easily. As might be expected, then, enzymes that *remove* acetyl groups—**histone deacetylases**—stabilize the nucleosomes and prevent transcription.

Histone methylation, the addition of methyl groups to histones by **histone methyltransferases**, can either activate or further repress transcription, depending on the amino acid being methylated and the presence of other methyl or acetyl groups in the vicinity (see Strahl and Allis 2000; Cosgrove et al. 2004). For instance, acetylation of the "tails" of H3 and H4 along with methylation of the lysine at position 4 of H3 (i.e., H3K4; remember that K is the abbreviation for lysine) is usually associated with actively transcribed chromatin. In contrast, a combined lack of acetylation of the H3 and H4 tails and methylation of the lysine in the ninth position of H3 (H3K9) is usually associated with highly repressed chromatin (Norma et al. 2001). Indeed, lysine methylations at H3K9, H327, and H4K20 are often associated with highly repressed chromatin. Figure 2.4 shows a schematic drawing of a nucleosome, with the histone H3 tail having on it some residues whose modification can regulate transcription.

As might be expected, if methyl groups at specific places on the histones repress transcription, then getting rid of these methyl moieties should permit transcription. This has been shown in the activation of those genes responsible for specifying the posterior halves of vertebrate bodies. These genes, called Hox genes, encode transcription factors that are critical in giving cells their identities along the anterior-posterior axis. In early development, Hox genes are repressed by H3K27 trimethylation (the lysine at position 27 having three methyl groups). However, in differentiated cells, a demethylase that is specific for H3K27me3 is recruited to these promoters and enables the gene to be transcribed (Agger et al. 2007; Lan et al. 2007).

The effects of methylation in controlling gene transcription are extensive. So far, we have documented transcriptional regulation by *histone* methylation. Later in this chapter we will discuss the exciting research on the control of transcription by *DNA* methylation.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

HISTONE REGULATION OF TRANSCRIPTIONAL ELONGATION
In addition to regulating the initiation of the transcriptional complex (i.e., getting RNA polymerase on the promoter), nucleosomes also appear to regulate the progression of RNA polymerase and the elongation of the mRNA. Indeed, recent evidence suggests that it is relatively common for RNA polymerase to be poised at the promoters, ready to go. For transcription to occur, these nucleosomes need to be modified, and it is possible that the acetylation of histone H3 at positions 9 and 14, coupled with the trimethylation of that histone at position 4, is critical for allowing elongation of the message (Guenther et al. 2007; Li et al. 2007).

Anatomy of the gene: Exons and introns

The second difference between prokaryotic and eukaryotic genes is that eukaryotic genes are not co-linear with their peptide products. Rather, the single nucleic acid strand of eukaryotic mRNA comes from noncontiguous regions on the chromosome. Between **exons**—the regions of DNA that code for a protein*—are intervening sequences called **introns** that have nothing whatsoever to do with the amino acid sequence of the protein. The structure of a typical eukaryotic gene can be illustrated by the human β-globin gene (Figure 2.5). This gene, which encodes part of the hemoglobin protein of the red blood cells, consists of the following elements:

- A **promoter region**, which is responsible for the binding of RNA polymerase and for the subsequent initiation of transcription. The promoter region of the human β-globin gene has three distinct units and extends from 95 to 26 base pairs before ("upstream from")† the transcription initiation site (i.e., from –95 to –26).
- The **transcription initiation site**, which for human β-globin is ACATTTG. This site is often called the **cap sequence** because it represents the 5′ end of the RNA, which will receive a "cap" of modified nucleotides soon after it is transcribed. The specific cap sequence varies among genes.
- The **translation initiation site**, ATG. This codon (which becomes AUG in mRNA) is located 50 base pairs after the transcription initiation site in the human β-globin gene (although this distance differs greatly among different genes). The sequence of 50 base pairs intervening between the initiation points of transcription and translation is the **5′ untranslated region**, often called the **5′ UTR** or **leader sequence.** The 5′ UTR can determine the rate at which translation is initiated.
- The first exon, which contains 90 base pairs coding for amino acids 1–30 of human β-globin protein.
- An intron containing 130 base pairs with no coding sequences for β-globin. However, the structure of this intron is important in enabling the RNA to be processed into mRNA and exit from the nucleus.
- An exon containing 222 base pairs coding for amino acids 31–104.
- A large intron—850 base pairs—having nothing to do with globin protein structure.
- An exon containing 126 base pairs coding for amino acids 105–146 of the protein.
- A **translation termination codon**, TAA. This codon becomes UAA in the mRNA. The ribosome dissociates at this codon, and the protein is released.
- A **3′ untranslated region** (**3′ UTR**) that, although transcribed, is not translated into protein. This region includes the sequence AATAAA, which is needed for **polyadenylation**, the insertion of a "tail" of some 200–300 adenylate residues on the RNA transcript, about 20 bases downstream of the AAUAAA sequence. This polyA tail (1) confers stability on the mRNA, (2) allows the mRNA to exit the nucleus, and (3) permits the mRNA to be translated into protein.
- A **transcription termination sequence**. Transcription continues beyond the AATAAA site for about 1000 nucleotides before being terminated.

The original transcription product is called **nuclear RNA (nRNA),** sometimes called *heterogeneous nuclear RNA* (hnRNA) or *pre-messenger RNA* (pre-mRNA). Nuclear RNA contains the cap sequence, the 5′ UTR, exons, introns, and the 3′ UTR (Figure 2.6). Both ends of these transcripts are modified before these RNAs leave the nucleus. A cap consisting of methylated guanosine is placed on the 5′ end of the RNA in opposite polarity to the RNA itself. This means there is no free 5′ phosphate group on the nRNA. The 5′ cap is necessary for the binding of mRNA to the ribosome and for subsequent translation (Shatkin 1976). The 3′ terminus is usually modified in the nucleus by the addition of a polyA tail. The adenylate residues in this tail are put together enzymatically and are added to the transcript; they are not part of the gene sequence. Both the 5′ and 3′ modifications may protect the mRNA from exonucleases that would otherwise digest it (Sheiness and Darnell 1973; Gedamu and Dixon 1978). The modifications thus stabilize the message and its precursor.

As the nRNA leaves the nucleus, its introns are removed and the remaining exons spliced together. In this way the coding regions of the mRNA—i.e., the exons—are brought together to form a single transcript, and this transcript is translated into a protein. The protein can be further modified to make it functional (see Figure 2.6).

*The term *exon* refers to a nucleotide sequence whose RNA "exits" the nucleus. It has taken on the functional definition of a protein-encoding nucleotide sequence. Leader sequences and 3′ UTR sequences are also derived from exons, even though they are not translated into protein.

†By convention, upstream, downstream, 5′, and 3′ directions are specified in relation to the RNA. Thus, the promoter is upstream of the gene, near its 5′ end.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

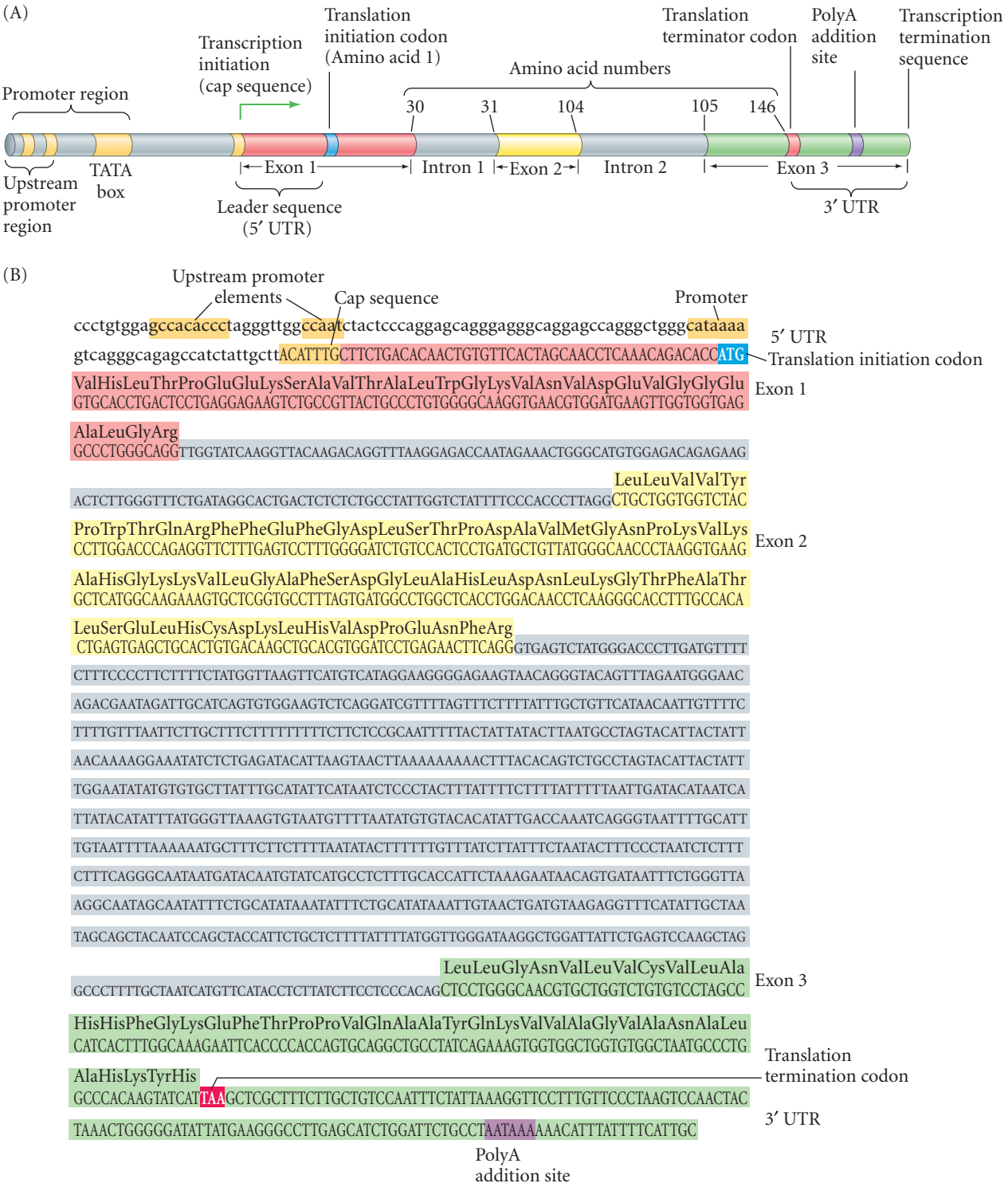


FIGURE 2.5 Nucleotide sequence of the human β-globin gene. (A) Schematic representation of the locations of the promoter region, transcription initiation site (cap sequence), 5′ untranslated region (leader sequence), exons, introns, and 3′ untranslated region. Exons are shown in color; the numbers flanking them indicate the amino acid positions each exon encodes in β-globin. (B) The nucleotide sequence shown from the 5′ end to the 3′ end of the RNA. The colors correspond to their diagrammatic representation in (A). The promoter sequences are boxed, as are the translation initiation and termination codes ATG and TAA. The large capital letters boxed in color are the bases of the exons, with the amino acids for which they code abbreviated above them. Smaller capital letters indicate the intron bases. The codons after the translation termination site exist in β-globin mRNA but are not translated into proteins. Within this group is the sequence thought to be needed for polyadenylation. By convention, only the RNA-like strand of the DNA double helix is shown. (B after Lawn et al. 1980.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

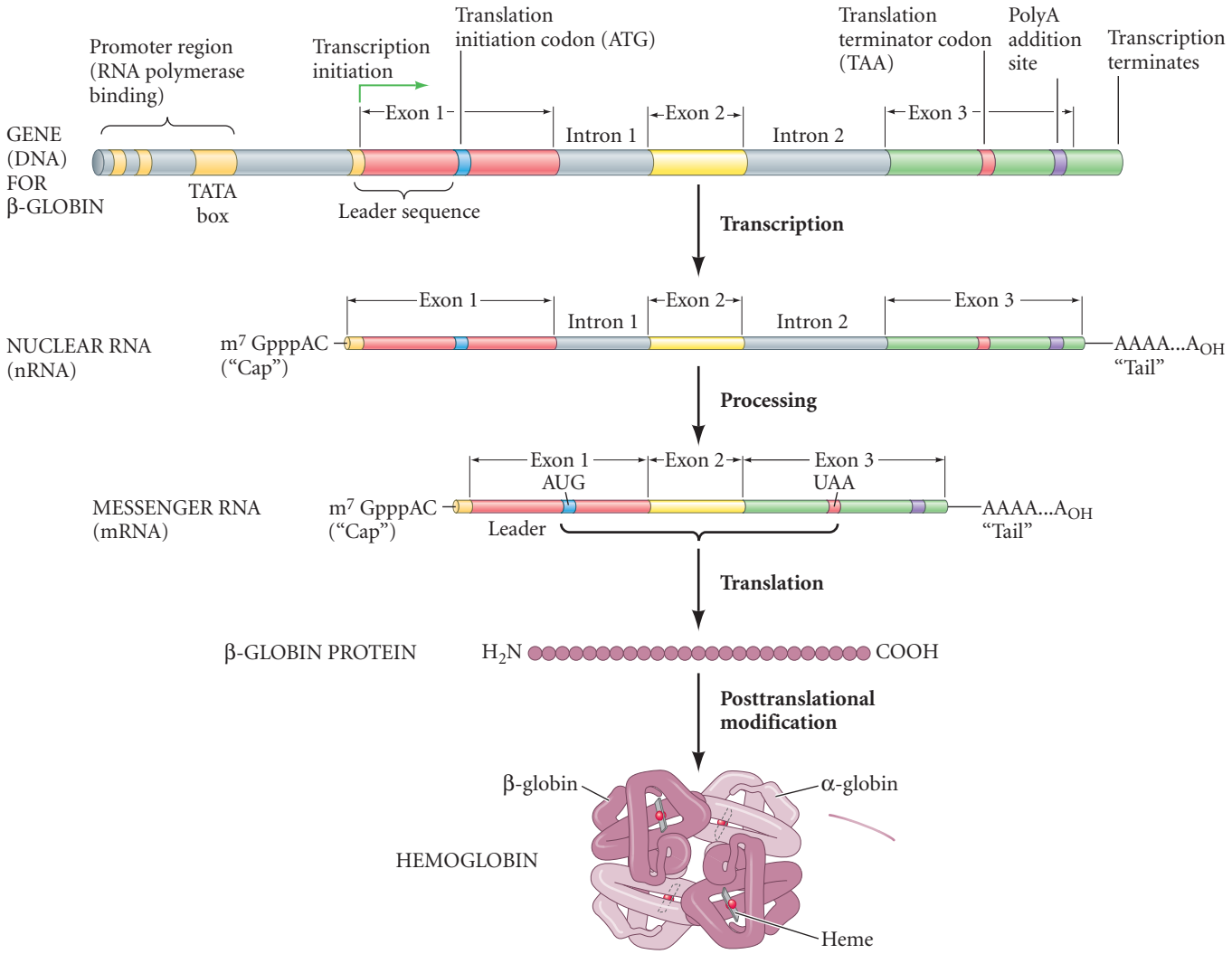


FIGURE 2.6 Summary of steps involved in the production of β-globin and hemoglobin. Transcription of the gene creates a nuclear RNA containing exons and introns, as well as the cap, tail, and 3′ and 5′ untranslated regions. Processing the nuclear RNA into messenger RNA removes the introns. Translation on ribosomes uses the mRNA to encode a protein. The protein is inactive until it is modified and complexed with α-globin and heme to become active hemoglobin (bottom).

Anatomy of the gene: Promoters and enhancers

In addition to the protein-encoding region of the gene, there are regulatory sequences that can be located on either end of the gene (or even within it). These sequences—the promoters and enhancers—are necessary for controlling where and when a particular gene is transcribed.

Promoters are the sites where RNA polymerase binds to the DNA to initiate transcription. Promoters of genes that synthesize messenger RNAs (i.e., genes that encode proteins*) are typically located immediately upstream from the site where the RNA polymerase initiates transcription. Most of these promoters contain the sequence TATA, to which RNA polymerase will be bound. This site, known as the **TATA box**, is usually about 30 base pairs upstream from the site where the first base is transcribed. Since this sequence will appear randomly in the genome at more places than just at promoter sites, other regions flanking it are also important. Many TATA box regions are flanked by

*There are several types of RNA that do *not* encode proteins. These include the ribosomal RNAs and transfer RNAs (which are used in protein synthesis) and the small nuclear RNAs (which are used in RNA processing). In addition, there are regulatory RNAs (such as the microRNAs that we will discuss later in this chapter), which are involved in regulating gene expression and are not translated into peptides.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

CpG islands, regions of DNA rich in those two nucleotides (Down and Hubbard 2002).

Eukaryotic RNA polymerases will not bind to the "naked" TATA sequence; they require the presence of additional proteins to place the polymerase properly on the promoter (Figure 2.7). Two of these are the **TATA-binding protein** (**TBP**), which forms a complex (TFIID) with other proteins to create a "saddle" upon which the RNA polymerase sits; and **TFIIB**, which recruits RNA polymerase to the TBP and positions it in such a manner that it can read the DNA codons (Kostrewa et al. 2009). Other proteins (TFIIA and TFIIH) stabilize the complex. In addition, auxiliary **transcription-associated factors** (**TAFs**) stabilize the RNA polymerase on the promoter and enable it to initiate transcription. These TAFs are bound by **upstream promoter elements** (sometimes called *proximal promoter sites*), which are DNA sequences near the TATA box and usually upstream from it. Eventually, TFIIH will phosphorylate the carboxy terminal of RNA polymerase, releasing it from the saddle so that it can transcribe the mRNA.

An **enhancer** is a DNA sequence that controls the efficiency and rate of transcription from a specific promoter. In other words, enhancers tell where and when a promoter can be used, and how much of the gene product to make. Enhancers bind specific **transcription factors**, proteins that activate the gene by (1) recruiting enzymes (such as histone acetyltransferases) that break up the nucleosomes in the area or (2) stabilizing the transcription initiation complex as described above.

Enhancers can activate only *cis*-linked promoters (i.e., promoters on the same chromosome*); therefore they are sometimes called ***cis*-regulatory elements**. However, because of DNA folding, enhancers can regulate genes at great distances (some as great as a million bases away) from the promoter (Visel et al. 2009). Moreover, enhancers do not need to be on the 5′ (upstream) side of the gene; they can be at the 3′ end, or even in the introns (Maniatis et al. 1987). The human β-globin gene has an enhancer in its 3′ UTR. This enhancer sequence is necessary for the temporal- and tissue-specific expression of the β-globin gene in adult red blood cell precursors (Trudel and Constantini 1987).

One of the principal methods of identifying enhancer sequences is to clone DNA sequences flanking the gene of interest and fuse them to **reporter genes** whose products are both readily identifiable and not usually made in the

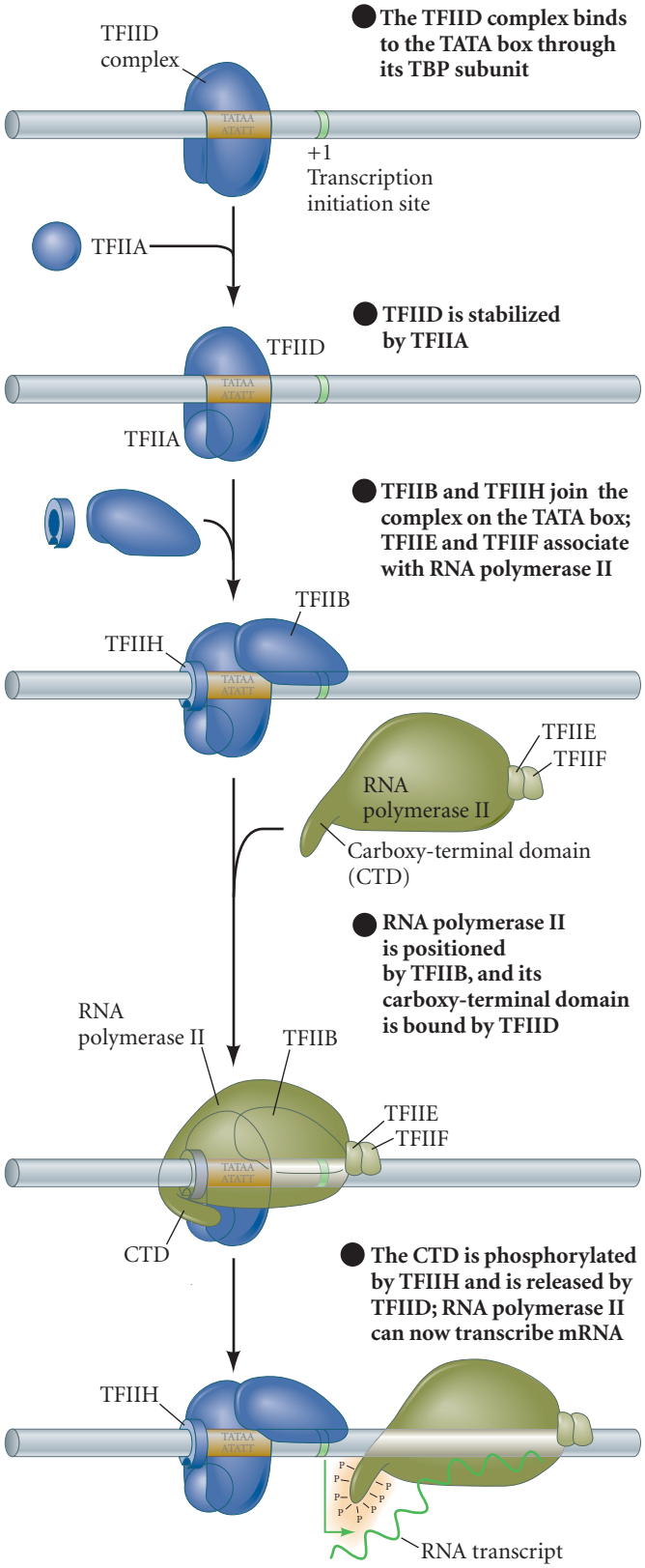
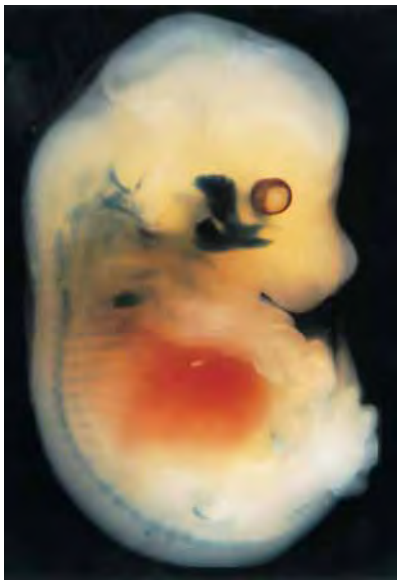


FIGURE 2.7 Formation of the active eukaryotic transcription preinitiation complex. The diagrams represent the formation of the complex that recruits and stabilizes RNA polymerase onto the promoter. *TF* stands for transcription factor; *II* indicates that the factor was first identified as being needed for RNA polymerase II (the RNA polymerase that transcribes protein-encoding genes); and the letters designate the particular active fraction from the phosphocellulose columns used to purify it.

*Cis- and *trans*-regulatory elements are so named by analogy with *E. coli* genetics and organic chemistry. There, *cis*-elements are regulatory elements that reside on the same strand of DNA (*cis*-, "on the same side as"), while *trans*-elements are those that could be supplied from another chromosome (*trans*-, "on the other side of"). The term *cis*-regulatory elements now refers to those DNA sequences that regulate a gene on the same stretch of DNA (i.e., the promoters and enhancers). *Trans*-regulatory factors are soluble molecules whose genes are located elsewhere in the genome and which bind to the *cis*-regulatory elements. They are usually transcription factors or microRNAs.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

(A)



(B)

FIGURE 2.8 The genetic elements regulating tissue-specific transcription can be identified by fusing reporter genes to suspected enhancer regions of the genes expressed in particular cell types. (A) The enhancer region of the gene for the muscle-specific protein Myf-5 is fused to a β-galactosidase reporter gene and incorporated into a mouse embryo. When stained for β-galactosidase activity (darkly staining region), the 13.5-day mouse embryo shows that the reporter gene is expressed in the eye muscles, facial muscles, forelimb muscles, neck muscles, and segmented myotomes (which give rise to the back musculature). (B) The *GFP* gene is fused to a lens crystallin gene in *Xenopus tropicalis*. The result is the expression of green fluorescent protein in the tadpole lens. (A courtesy of A. Patapoutian and B. Wold; B from Offield et al. 2000, photograph courtesy of R. Grainger.)

organism being studied. Researchers can insert constructs of possible enhancers and reporter genes into embryos and then monitor the expression of the reporter gene. If the sequence contains an enhancer, the reporter gene should become active at particular times and places. For instance, the *E. coli* gene for β-galactosidase (the ***lacZ* gene**) can be used as a reporter gene and fused to (1) a promoter that can be activated in any cell and (2) an enhancer that normally directs the expression of a particular mouse gene in muscles. When the resulting transgene is injected into a newly fertilized mouse egg and becomes incorporated into its DNA, β-galactosidase will be expressed in the mouse muscle cells. By staining for the presence of β-galactosidase, the expression pattern of that muscle-specific gene can be seen (Figure 2.8A).

Similarly, a sequence flanking a lens crystallin protein in *Xenopus* was shown to be an enhancer. When this sequence was fused to a reporter gene for green fluorescent protein (see Figure 1.17), GFP was expressed only in the lens (Figure 2.8B; Offield et al. 2000). GFP reporter genes are very useful because they can be monitored in live embryos and because the changes in gene expression can be seen in single cells.

ENHANCER MODULARITY The enhancer sequences on the DNA are the same in every cell type; what differs is the combination of transcription factor proteins the enhancers bind. Once bound to enhancers, transcription factors are able to enhance or suppress the ability of RNA polymerase to initiate transcription. Enhancers can bind several transcription factors, and it is the specific *combination* of transcription factors present that allows a gene to be active in a particular cell type. That is, the same transcription factor, in conjunction with different other factors, will activate different promoters in different cells. Moreover, the same gene can have several enhancers, with each enhancer binding transcription factors that enable that same gene to be expressed in different cell types.

Figure 2.9 illustrates this phenomenon for expression of the mouse *Pax6* gene in the cornea and pancreas. The mouse *Pax6* gene (which is expressed in the lens and retina of the eye, in the neural tube, and in the pancreas) has several enhancers (Figure 2.9A). The 5′ regulatory regions of the mouse *Pax6* gene were discovered by taking regions from its 5′ flanking sequence and introns and fusing them to a *lacZ* reporter gene. Each of these transgenes was then microinjected into newly fertilized mouse pronuclei, and the resulting embryos were stained for β-galactosidase (Figure 2.9B; Kammandel et al. 1998; Williams et al. 1998). Analysis of the results revealed that the enhancer farthest upstream from the promoter contains the regions necessary for *Pax6* expression in the pancreas, while a second enhancer activates *Pax6* expression in surface ectoderm (lens, cornea, and conjunctiva). A third enhancer resides in the leader sequence; it contains the sequences that direct *Pax6* expression in the neural tube. A fourth enhancer sequence, located in an intron shortly downstream of the translation initiation site, determines the expression of *Pax6* in the retina. The *Pax6* gene illustrates the principle of enhancer modularity, wherein having multiple, separate enhancers allows a protein to be expressed in several different tissues while not being expressed at all in others.

COMBINATORIAL ASSOCIATION While enhancers are modular *between* enhancers, there are co-dependent units *within* each enhancer. Enhancers contain regions of DNA that

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

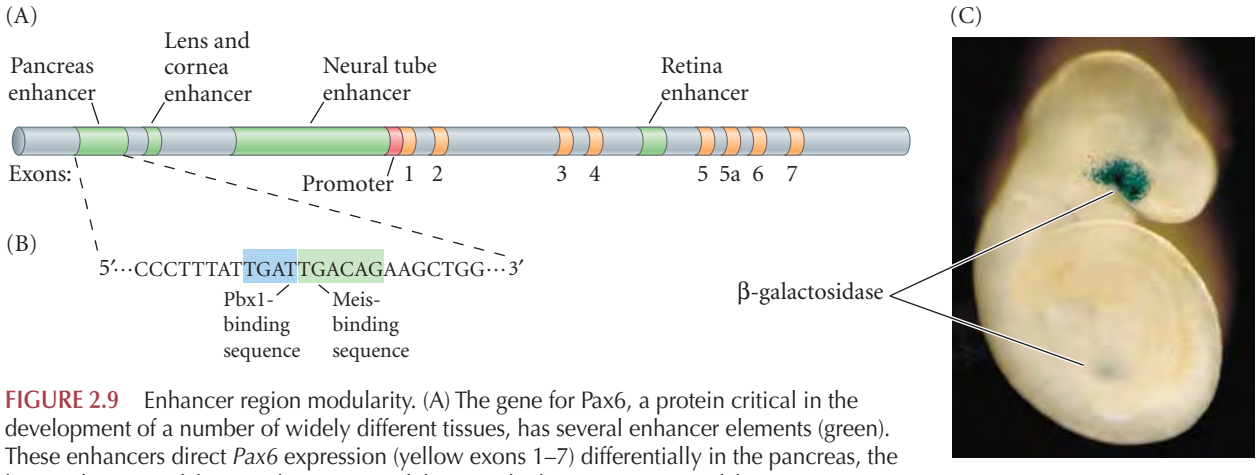


FIGURE 2.9 Enhancer region modularity. (A) The gene for Pax6, a protein critical in the development of a number of widely different tissues, has several enhancer elements (green). These enhancers direct *Pax6* expression (yellow exons 1–7) differentially in the pancreas, the lens and cornea of the eye, the retina, and the neural tube. (B) A portion of the DNA sequence of the pancreas-specific enhancer element. This sequence has binding sites for the Pbx1 and Meis transcription factors; both must be present in order to activate the *Pax6* gene in the pancreas.(C) When the gene for bacterial β-galactosidase is fused to *Pax6* enhancers for expression in the pancreas and the lens/cornea, this enzyme (which is easily stained) can be seen in those tissues. (C from Williams et al. 1998, courtesy of R. A. Lang.)

bind transcription factors, and it is this combination of transcription factors that activates the gene. For instance, the pancreas-specific enhancer of the *Pax6* gene has binding sites for the Pbx1 and Meis transcription factors (see Figure 2.9A). Both need to be present in order for the enhancer to activate *Pax6* in the pancreas cells (Zang et al. 2006).

Moreover, the product of the *Pax6* gene encodes a transcription factor that works in combinatorial partnerships with other transcription factors. Figure 2.10 shows two gene regulatory regions that bind Pax6. The first is that of the chick δ1 lens *crystallin* gene (**Figure 2.10A**; Cvekl and Piatigorsky 1996; Muta et al. 2002). This gene encodes crystallin, a lens protein that is transparent and allows light to reach the retina. A promoter within the *crystallin* gene contains a site for TBP binding, and an upstream promoter element that binds Sp1 (a general transcriptional activator found in all cells). The gene also has an enhancer in its third intron that controls the time and place of crystallin expression. This enhancer has two Pax6-binding sites. The Pax6 protein works with the Sox2 and L-Maf transcription factors to activate the *crystallin* gene only in those head cells that are going to become lens. As we will see in Chapter 10, this involves the cell being head ectoderm (which has Pax6), being in the region of the ectoderm likely to form eyes (L-Maf), and being in contact with the future retinal cells (which induce Sox2 expression; Kamachi et al. 1998).

Meanwhile, another set of regulatory regions that use Pax6 are the enhancers regulating the transcription of the genes for insulin, glucagon, and somatostatin in the pancreas (**Figure 2.10B**). Here, Pax6 is also essential for gene expression, and it works in cooperation with other transcription factors such as Pdx1 (specific for the pancreatic region of the endoderm) and Pbx1 (Andersen et al. 1999; Hussain and Habener 1999). In the absence of Pax6 (as in the homozygous *small eye* mutation in mice and rats), the endocrine cells of the pancreas do not develop properly and the production of hormones by those cells is deficient (Sander et al. 1997; Zhang et al. 2002).

There are other genes that are activated by Pax6 binding, and one of them is the *Pax6* gene itself. Pax6 protein can bind to a *cis*-regulatory element of the *Pax6* gene (Plaza et al. 1993). This means that once the *Pax6* gene is turned on, it will continue to be expressed, even if the signal that originally activated it is no longer given.

One can see that the genes for specific proteins use numerous transcription factors in various combinations. Thus, *enhancers are modular* (such that the *Pax6* gene is expressed in the eye, pancreas, and nervous system, as shown in Figure 2.9); but *within each cis-regulatory module, transcription factors work in a combinatorial fashion* (such that Pax6, L-Maf, and Sox2 are all needed for the transcription of crystallin in the lens). The combinatorial association of transcription factors on enhancers leads to the spatiotemporal output of any particular gene (see Davidson 2006; Zinzen et al. 2009).

Transcription factor function

Natalie Angier (1992) has written, "A series of new discoveries suggests that DNA is more like a certain type of politician, surrounded by a flock of protein handlers and advisers that must vigorously massage it, twist it, and on occasion, reinvent it before the grand blueprint of the body can make any sense at all." These "handlers and advisers" are the transcription factors. These factors can be grouped together in families based on similarities in structure (**Table 2.1**). The transcription factors within such a family share a common framework in their DNA-binding sites, and slight

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

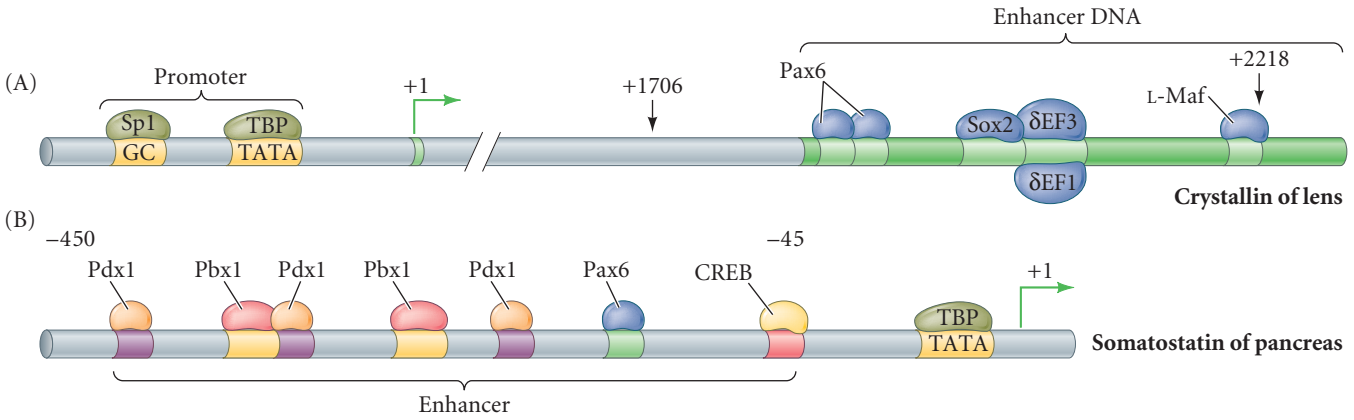


FIGURE 2.10 Modular transcriptional regulatory regions using Pax6 as an activator. (A) Promoter and enhancer of the chick δ1 lens *crystallin* gene. Pax6 interacts with two other transcription factors, Sox2 and L-Maf, to activate this gene. The protein δEF3 binds factors that permit this interaction; δEF1 binds factors that inhibit it. (B) Promoter and enhancer of the rat somatostatin gene. Pax6 activates this gene by cooperating with the Pbx1 and Pdx1 transcription factors. (A after Cvekl and Piatigorsky 1996; B after Andersen et al. 1999.)

differences in the amino acids at the binding site can cause the binding site to recognize different DNA sequences.

As we have already seen, enhancers function by binding transcription factors, and each enhancer can have binding sites for several transcription factors. Transcription factors bind to the enhancer DNA with one part of the protein and use other sites on the protein to interact with one another to recruit histone-modifying enzymes.

For example, the association of the Pax6, Sox2, and L-Maf transcription factors in lens cells recruits a histone acetyltransferase that can transfer acetyl groups to the histones and dissociate the nucleosomes in that area (Yang et al. 2006). Similarly, when MITF, a transcription factor essential for ear development and pigment production, binds to its specific DNA sequence, it also binds a (different) histone acetyltransferase that also facilitates the dissociation of nucleosomes (Ogryzko et al. 1996; Price et al. 1998). And the Pax7 transcription factor that activates muscle-specific genes binds to the enhancer region of these genes within the muscle precursor cells. Pax7 then recruits a histone methyltransferase that methylates the lysine in the fourth position of histone H3 (H3K4), resulting in the trimethylation of this lysine and the activation of transcription (McKinnell et al. 2008). The displacement of nucleosomes along the DNA makes it possible for the transcription fac-

TABLE 2.1 Some major transcription factor families and subfamilies

Family	Representative transcription factors	Some functions
Homeodomain:		
Hox	Hoxa1, Hoxb2, etc.	Axis formation
POU	Pit1, Unc-86, Oct-2	Pituitary development; neural fate
LIM	Lim1, Forkhead	Head development
Pax	Pax1, 2, 3, 6, etc.	Neural specification; eye development
Basic helix-loop-helix (bHLH)	MyoD, MITF, daughterless	Muscle and nerve specification; *Drosophila* sex determination; pigmentation
Basic leucine zipper (bZip)	C/EBP, AP1	Liver differentiation; fat cell specification
Zinc finger:		
Standard	WT1, Krüppel, Engrailed	Kidney, gonad, and macrophage development; *Drosophila* segmentation
Nuclear hormone receptors	Glucocorticoid receptor, estrogen receptor, testosterone receptor, retinoic acid receptors	Secondary sex determination; craniofacial development; limb development
Sry-Sox	Sry, SoxD, Sox2	Bend DNA; mammalian primary sex determination; ectoderm differentiation

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

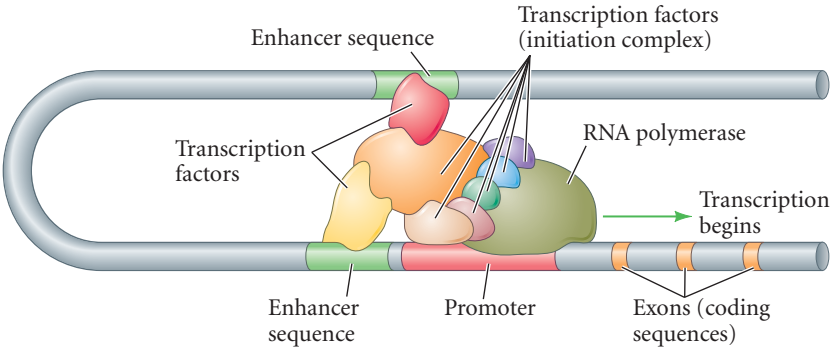


FIGURE 2.11 RNA polymerase is stabilized on the promoter site of the DNA by transcription factors recruited by the enhancers. The TATA sequence at the promoter binds a protein that serves as a "saddle" for RNA polymerase. However, RNA polymerase would not remain bound long enough to initiate transcription were it not for the stabilization by the transcription factors.

tors to find their binding sites (Adkins et al. 2004; Li et al. 2007).

In addition to recruiting nucleosome modifying enzymes, transcription factors can also work by stabilizing the transcription preinitiation complex that enables RNA polymerase to bind to the promoter (Figure 2.11). For instance, MyoD, a transcription factor that is critical for muscle cell development (see Chapter 11), stabilizes TFIIB, which supports RNA polymerase at the promoter site (Heller and Bengal 1998). Indeed, MyoD plays several roles in activating gene expression, since it also can bind histone acetyltransferases that initiate nucleosome remodeling and dissociation (Cao et al. 2006).

One of the important consequences of the combinatorial association of transcription factors is **coordinated gene expression**. The simultaneous expression of many cell-specific genes can be explained by the binding of transcription factors by the enhancer elements. For example, many genes that are specifically activated in the lens contain an enhancer that binds Pax6. This means that all the other transcription factors might be assembled at the enhancer, but until Pax6 binds, they cannot activate the gene. Similarly, many of the co-expressed muscle-specific genes contain enhancers that bind the MEF2 transcription factor; and the enhancers on genes encoding pigment-producing enzymes bind MITF (see Davidson 2006).

TRANSCRIPTION FACTOR DOMAINS Transcription factors have three major domains. The first is a **DNA-binding domain** that recognizes a particular DNA sequence in the enhancer. Figure 2.12 shows a model of such a domain in the Pax6 protein described earlier (see Figure 2.9). The second is a ***trans*-activating domain** that activates or suppresses the transcription of the gene whose promoter or enhancer it has bound. Usually, this *trans*-activating domain enables the transcription factor to interact with the proteins involved in binding RNA polymerase (such as TFIIB or TFIIE; see Sauer et al. 1995) or with enzymes that modify histones. In addition, there may be a **protein-protein interaction domain** that allows the transcription factor's activity to be modulated by TAFs or other transcription factors.

MITF, a transcription factor essential for ear development and pigment production, has a protein-protein interaction domain that enables it to dimerize with another MITF protein (Ferré-D'Amaré et al. 1993). The resulting homodimer (i.e., two identical protein molecules bound

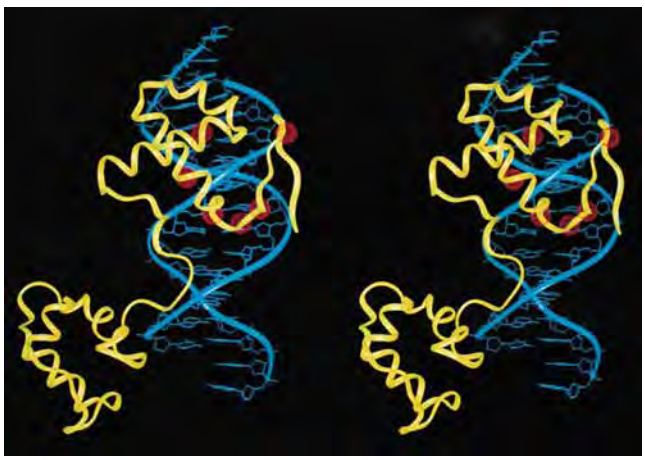


FIGURE 2.12 Stereoscopic model of Pax6 protein binding to its enhancer element in DNA. The DNA-binding region of Pax6 is shown in yellow; the DNA double helix is blue. Red dots indicate the sites of loss-of-function mutations in the *Pax6* gene that give rise to nonfunctional Pax6 proteins. It is worth trying to cross your eyes to see the central three-dimensional figure. (From Xu et al. 1995; photograph courtesy of S. O. Pääbo.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

SIDELIGHTS & SPECULATIONS

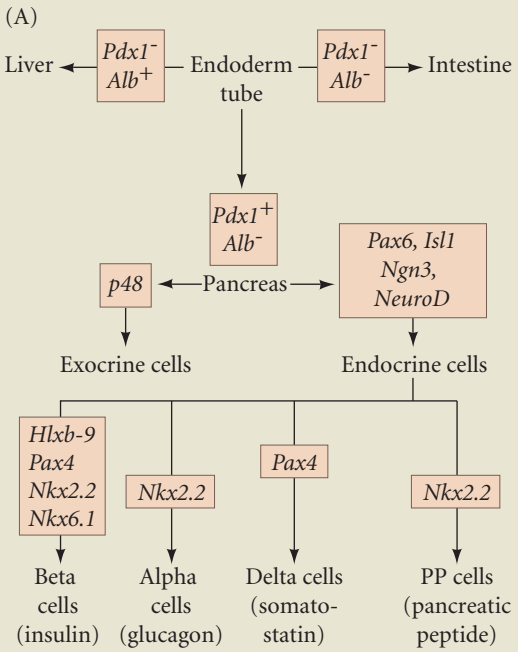
Reprogramming Cells: Changing Cell Differentiation through Embryonic Transcription Factors

The importance and power of transcription factors were elegantly demonstrated when Zhou and colleagues (2008) used three transcription factors to convert *exocrine* pancreatic cells (which make amylase, chymotrypsin, and other digestive enzymes) into insulin-secreting *endocrine* pancreatic β cells. The researchers infected the pancreases of living 2-month-old mice with harmless viruses containing the genes for three transcription factors: Pdx1, Ngn3, and Mafa.

The Pdx1 protein stimulates the outgrowth of the digestive tube that results in the pancreatic buds. This protein is found throughout the pancreas and is critical in specifying that organ's endocrine cells, as well as in activating genes that encode endocrine proteins (see Figure 2.10B). Ngn3 is a transcription factor found in endocrine, but not exocrine, pancreatic cells. Mafa, a transcription factor regulated by glucose levels, is found only in pancreatic β cells (i.e., those cells that make insulin) and can activate transcription of the insulin gene.

Pdx1, Ngn3, and Mafa activate other transcription factors that work in concert to turn a pancreatic endodermal cell into an insulin-secreting β cell. Zhou and colleagues found that, of all the transcription factor genes tested, these three were the only ones that were crucial for the conversion (Figure 2.13). Converted pancreas cells looked identical to normal β cells, and like normal β cells, they were capable of secreting both insulin and a blood vessel-inducing factor. The converted cells retained their new properties for months. Moreover, mice whose normal β cells were destroyed by a particular drug developed diabetes, just like the Type 1 diabetes seen when human adults lose β cells. This diabetes could be cured by injecting the mice with viruses containing the three transcription factors. When this was done, about 20% of the exocrine pancreatic cells became β cells and secreted insulin.

This study opens the door to an entire new field of regenerative medicine, illustrating the possibilities of changing one adult cell type into another by using the transcription factors that had made the new cell type in the embryo.



(B)

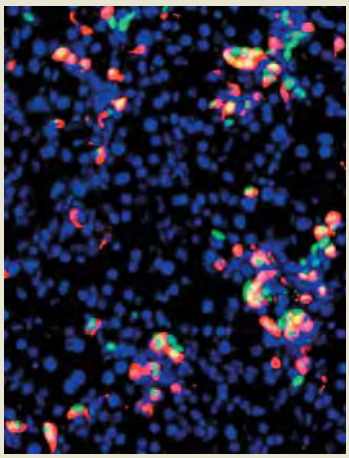


Figure 2.13 Pancreatic lineage and transcription factors. (A) Pdx1 protein is expressed in pancreatic lineages. Several transcription factors, including Ngn3, distinguish the endocrine lineage. Several other transcription factors, including Mafa, are found in the β cells that produce insulin. (B) New pancreatic β cells arise in adult mouse pancreas in vivo after viral delivery of three transcription factors (Ngn3, Pdx1, and Mafa). Virally infected exocrine cells are detected by their expression of nuclear green fluorescent protein. Newly induced β cells are recognized by insulin staining (red). Their overlap produces yellow. The nuclei of all pancreatic cells are stained blue. (B courtesy of D. Melton.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

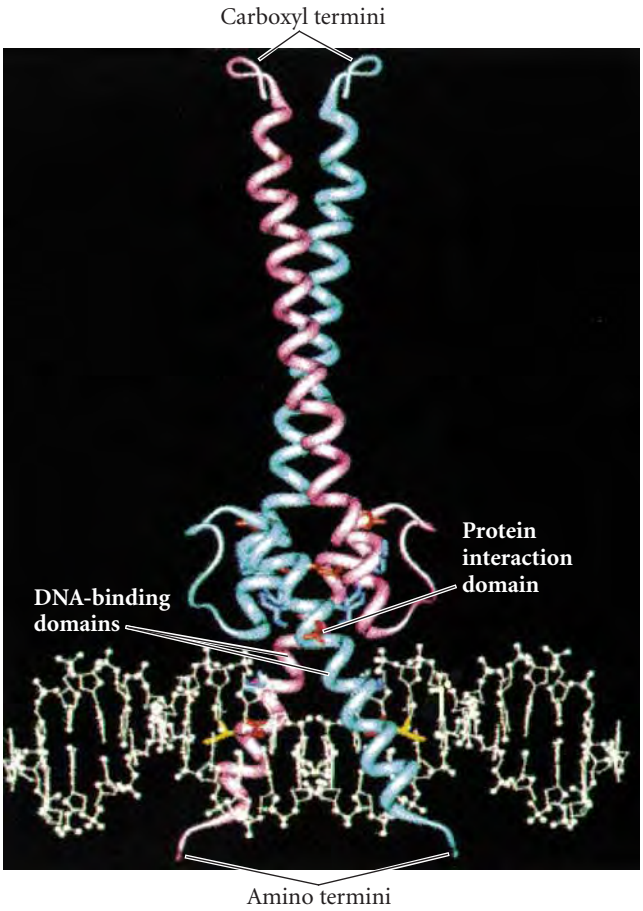


FIGURE 2.14 Three-dimensional model of the homodimeric transcription factor MITF (one protein shown in red, the other in blue) binding to a promoter element in DNA (white). The amino termini are located at the bottom of the figure and form the DNA-binding domain that recognizes an 11-base-pair sequence of DNA having the core sequence CATGTG. The protein-protein interaction domain is located immediately above. MITF has the basic helix-loop-helix structure found in many transcription factors. The carboxyl end of the molecule is thought to be the *trans*-activating domain that binds the p300/CBP co-activator protein. (From Steingrímsson et al. 1994; photograph courtesy of N. Jenkins.)

together) is a functional protein that can bind to DNA and activate the transcription of certain genes (Figure 2.14). The DNA-binding domain of MITF is close to the amino-terminal end of the protein and contains numerous basic amino acids that make contact with the DNA (Hemesath et al. 1994; Steingrímsson et al. 1994). This assignment was confirmed by the discovery of various human and mouse mutations that map within the DNA-binding site for MITF and which prevent the attachment of the MITF protein to the DNA. Sequences for MITF binding have been found in the regulatory regions of genes encoding three pigment-cell-specific enzymes of the tyrosinase family (Bentley et al. 1994; Yasumoto et al. 1997). Without MITF, these proteins are not synthesized properly, and melanin pigment is not made. These *cis*-regulatory regions all contain the same 11-base-pair sequence, including the core sequence (CATGTG) that is recognized by MITF.

The third functional region of MITF is its *trans*-activating domain. This domain includes a long stretch of amino acids in the center of the protein. When the MITF dimer is bound to its target sequence in the enhancer, the *trans*-activating region is able to bind a TAF, p300/CBP. The p300/CPB protein is a histone acetyltransferase enzyme that can transfer acetyl groups to each histone in the nucleosomes (Ogryzko et al. 1996; Price et al. 1998). Acetylation of the nucleosomes destabilizes them and allows the genes for pigment-forming enzymes to be expressed.

EPIGENETIC MEMORY: KEEPING THE RIGHT GENES ON OR OFF

The modifications of histones can also signal the recruitment of the proteins that can retain the memory of transcriptional state from generation to generation through mitosis. These are the proteins of the Trithorax and Polycomb families. When bound to the nucleosomes of active genes, **Trithorax** proteins keep these genes active, whereas **Polycomb** proteins, which bind to condensed nucleosomes, keep the genes in an inactive state.

The Polycomb proteins fall into two categories that act sequentially in repression. The first set has histone methyltransferase activities that methylate lysines H3K27 and H3K9 to repress gene activity. In many organisms, this repressive state is stabilized by the activity of a second set of Polycomb factors, which bind to the methylated tails of histone 3 and keep the methylation active and also methylate adjacent nucleosomes, thereby forming tightly packed repressive complexes (Grossniklaus and Paro 2007; Margueron et al. 2009).

The Trithorax proteins help retain the memory of activation; they act to counter the effect of the Polycomb proteins. Trithorax proteins can modify the nucleosomes or alter their positions on the chromatin, allowing transcription factors to bind to the DNA previously covered by them. Other Trithorax proteins keep the H3K4 lysine trimethylated (preventing its demethylation into a dimethylated, repressive, state; Tan et al. 2008).

PIONEER TRANSCRIPTION FACTORS: BREAKING THE SILENCE

Finding a promoter is not easy, because the DNA is usually so wound up that the promoter sites are not accessible. Indeed, more than *6 feet* of DNA is packaged into chromosomes of each human cell nucleus (Schones and Zhao 2008).

How can a transcription factor find its binding site, given that the enhancer might be covered by nucleosomes? Several studies have identified certain transcription factors that penetrate repressed chromatin and bind to their enhancer DNA sequences (Cirillo et al. 2002;

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

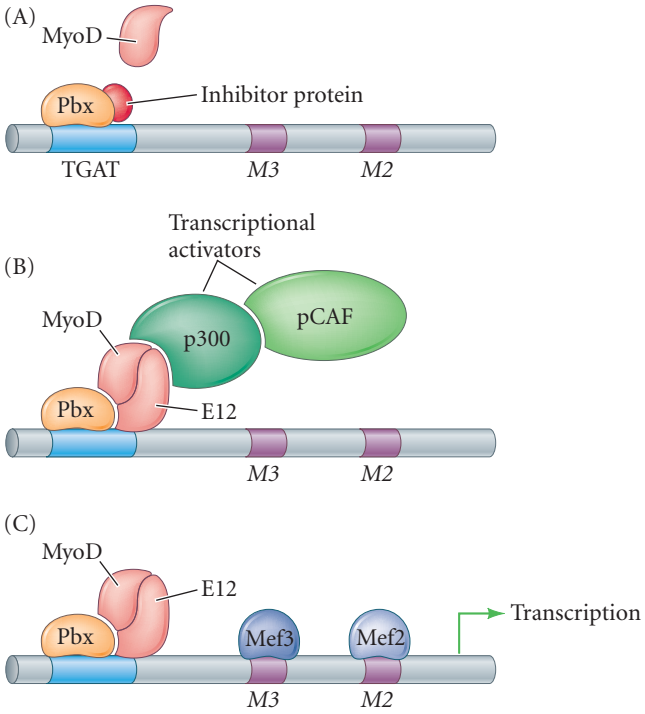


FIGURE 2.15 Model for the role of the "pioneer" transcription factor Pbx in aligning the muscle-specific transcription factor MyoD on DNA. (A) Pbx protein recognizes its DNA binding site (TGAT), even within nucleosome-rich chromatin. Pbx probably binds to transcriptional inhibitors. (B) MyoD, complexed with its E12 cofactor, is able to bind to Pbx, replacing the transcriptional inhibitors. MyoD then binds to its recognition element on the DNA. (C) The MyoD/E12 complex can then recruit the histone acetyltransferases and nucleosome remodeling compounds that make the chromatin accessible to other transcription factors (Mef3 and Mef2) and to RNA polymerase. (After Tapscott 2005.)

Berkes et al. 2004). They have been called "pioneer" transcription factors, and they appear to be critical in establishing certain cell lineages. One of these transcription factors is FoxA1, which binds to certain enhancers and opens up the chromatin to allow other transcription factors access to the promoter (Lupien et al. 2008). FoxA proteins remain bound to the DNA during mitosis, providing a mechanism to re-establish normal transcription in presumptive liver cells (Zaret et al. 2008). Another pioneer transcription factor is the Pax7 protein mentioned above. It activates muscle-specific gene transcription in a population of muscle stem cells by binding to its DNA recognition sequence and being stabilized there by dimethylated H3K4 on the nucleosomes. It then recruits the histone methyltransferase that converts the dimethylated H3K4 into the trimethylated H3K4 associated with active transcription (McKinnell et al. 2008).

Another pioneer transcription factor in muscle development is Pbx. Members of the Pbx family are made in every cell, and they are able to find their appropriate sites even in highly compacted chromatin. Pbx appears to be used as a "molecular beacon" for another muscle-determining transcription factor, MyoD (mentioned earlier). MyoD is critical for initiating muscle development in the embryo, activating hundreds of genes that are involved with establishing the muscle phenotype. However, MyoD is not able to bind to DNA without the help of Pbx proteins, which bind to DNA elements adjacent to the DNA sequence recognized by MyoD (Figure 2.15A). Berkes and colleagues (2004) have shown that MyoD (when complexed with another transcription factor, E12) can bind to the Pbx protein and align itself on its target DNA sequence (Figure 2.15B). Once bound there, the E12 protein recruits histone acetyltransferases and nucleosome remodeling complexes to open up the DNA on those genes (Figure 2.15C).

SILENCERS **Silencers** are DNA regulatory elements that actively repress the transcription of a particular gene. They can be viewed as "negative enhancers." For instance, in the mouse, there is a DNA sequence that prevents a promoter's activation in any tissue *except* neurons. This sequence, given the name **neural restrictive silencer element** (**NRSE**), has been found in several mouse genes whose expression is limited to the nervous system: those encoding synapsin I, sodium channel type II, brain-derived neurotrophic factor, Ng-CAM, and L1. The protein that binds to the NRSE is a zinc finger transcription factor called **neural restrictive silencer factor** (**NRSF**). (It is also called REST). NRSF appears to be expressed in every cell that is *not* a mature neuron (Chong et al. 1995; Schoenherr and Anderson 1995).

To test the hypothesis that the NRSE sequence is necessary for the normal repression of neural genes in non-neural cells, transgenes were made by fusing a β-galactosidase (*lacZ*) gene with part of the *L1* neural cell adhesion gene. (L1 is a protein whose function is critical for brain development, as we will see in later chapters.) In one case, the *L1* gene, from its promoter through the fourth exon, was fused to the *lacZ* sequence. A second transgene was made just like the first, except that the NRSE was deleted from the *L1* promoter. The two transgenes were separately inserted into the pronuclei of fertilized oocytes, and the resulting transgenic mice were analyzed for β-galactosidase expression (Kallunki et al. 1995, 1997). In embryos receiving the complete transgene (which included the NSRE), expression was seen only in the nervous system (Figure 2.16A). In mice whose transgene lacked the NRSE sequence, however, expression was seen in the heart, the limb mesenchyme and limb ectoderm, the kidney mesoderm, the ventral body wall, and the cephalic mesenchyme (Figure 2.16B).

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

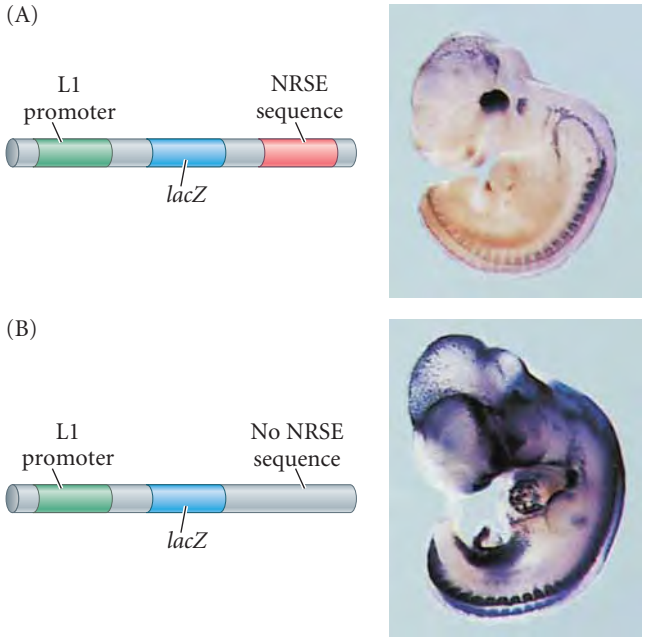


FIGURE 2.16 Silencers. Analysis of β-galactosidase staining patterns in 11.5-day embryonic mice. (A) Embryo containing a transgene composed of the L1 promoter, a portion of the *L1* gene, and a *lacZ* gene fused to the second exon (which contains the NRSE region). (B) Embryo containing a similar transgene but lacking the NRSE sequence. The dark areas reveal the presence of β-galactosidase (the *lacZ* product). (Photographs from Kallunki et al. 1997.)

DNA Methylation and the Control of Transcription

How does a pattern of gene transcription become stable? How can a lens cell continue to remain a lens cell and not activate muscle-specific genes? How can cells undergo rounds of mitosis and still maintain their differentiated characteristics? The answer appears to be **DNA methylation**. We have already discussed *histone* methylation and its importance for transcription. Now we look at how the *DNA itself* can be methylated to regulate transcription. Generally speaking, the promoters of inactive genes become methylated at certain cytosine residues, and the resulting methylcytosine stabilizes nucleosomes and prevents transcription factors from binding.

It is often assumed that a gene contains exactly the same nucleotides whether it is active or inactive; that is, a β-globin gene that is activated in a red blood cell precursor has the same nucleotides as the inactive β-globin gene in a fibroblast or retinal cell of the same animal. However, it turns out there is in fact a subtle difference. In 1948, R. D. Hotchkiss discovered a "fifth base" in DNA, 5-methylcytosine. In vertebrates, this base is made enzymatically after DNA is replicated. At this time, about 5% of the cytosines in mammalian DNA are converted to 5-methylcytosine (Figure 2.17A). This conversion can occur only when the cytosine residue is followed by a guanosine—in other words, at a CpG sequence (as we will soon see, this restriction is important). Numerous studies have shown that the degree to which the cytosines of a gene are methylated can control the level of the gene's transcription. Cytosine methylation appears to be a major mechanism of transcriptional regulation among vertebrates; however, some other species (*Drosophila* and nematodes among them) do not methylate their DNA.

In vertebrates, the presence of methylated cytosines in a gene's promoter correlates with the repression of transcription from that gene. In developing human and chick red blood cells, for example, the DNA of the globin gene promoters is almost completely unmethylated, whereas the same promoters are highly methylated in cells that do not produce globins. Moreover, the methylation pattern changes during development (Figure 2.17B). The cells that produce hemoglobin in the human embryo have unmethylated promoters in the genes encoding the ε-globins ("embryonic globin chains") of embryonic hemoglobin. These promoters become methylated in the fetal tissue, as the genes for fetal-specific γ-globin (rather than the embryonic chains) become activated (van der Ploeg and Flavell 1980; Groudine and Weintraub 1981; Mavilio et al. 1983).

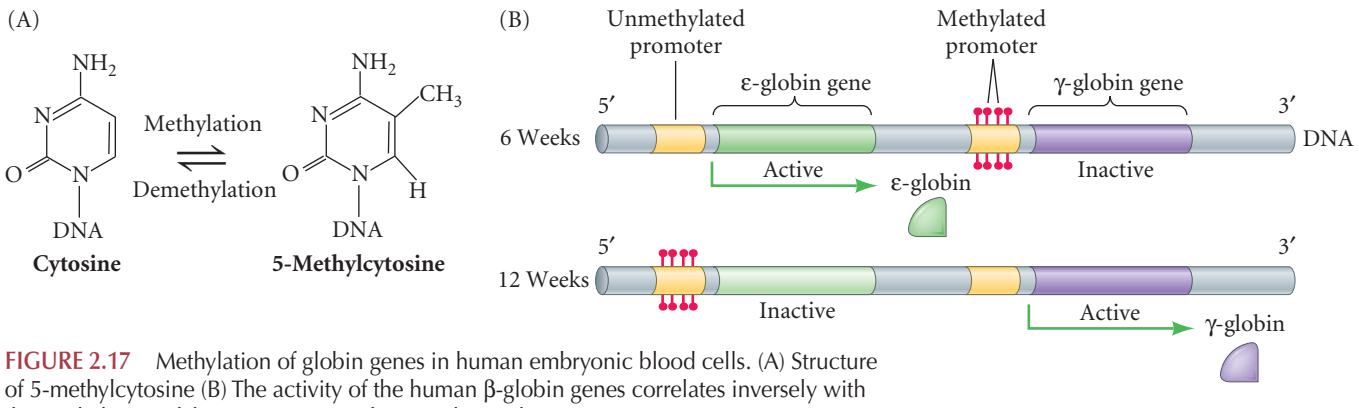


FIGURE 2.17 Methylation of globin genes in human embryonic blood cells. (A) Structure of 5-methylcytosine (B) The activity of the human β-globin genes correlates inversely with the methylation of their promoters. (After Mavilio et al. 1983.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

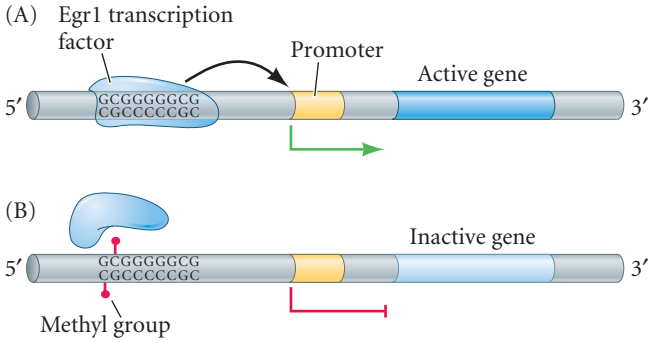


FIGURE 2.18 DNA methylation can block transcription by preventing transcription factors from binding to the enhancer region. (A) The Egr1 transcription factor can bind to specific DNA sequences such as 5′…GCGGGGGCG…3′, helping to activate transcription of those genes. (B) If the first cytosine residue is methylated, however, Egr1 will not bind and the gene will remain repressed. (After Weaver et al. 2005.)

Similarly, when fetal globin gives way to adult (β) globin, promoters of the fetal (γ) globin genes become methylated.

Mechanisms by which DNA methylation blocks transcription

DNA methylation appears to act in two ways to repress gene expression. First, it can block the binding of transcription factors to enhancers. Several transcription factors can bind to a particular sequence of unmethylated DNA, but they cannot bind to that DNA if one of its cytosines is methylated (Figure 2.18). Second, a methylated cytosine can recruit the binding of proteins that facilitate the methylation or deacetylation of histones, thereby stabilizing the nucleosomes. For instance, methylated cytosines in DNA can bind particular proteins such as MeCP2. Once connected to a methylated cytosine, MeCP2 binds to histone deacetylases and histone methyltransferases, which, respectively, remove acetyl groups (Figure 2.19A) and add methyl groups (Figure 2.19B) on the histones. As a result, the nucleosomes form tight complexes with the DNA and don't allow other transcription factors and RNA polymerases to find the genes. Other proteins, such as HP1 and histone H1, will bind and aggregate methylated histones (Fuks 2005; Rupp and Becker 2005). In this way, repressed chromatin becomes associated with regions where there are methylated cytosines.

Inheritance and stabilization of DNA methylation patterns

Another enzyme recruited to the chromatin by MeCP2 is DNA methyltransferase-3 (Dnmt3). This enzyme methylates previously unmethylated cytosines on the DNA. In this way, a relatively large region can be repressed. The newly established methylation pattern is then transmitted to the next generation by DNA methyltransferase-1 (Dnmt1). This enzyme recognizes methyl cytosines on one

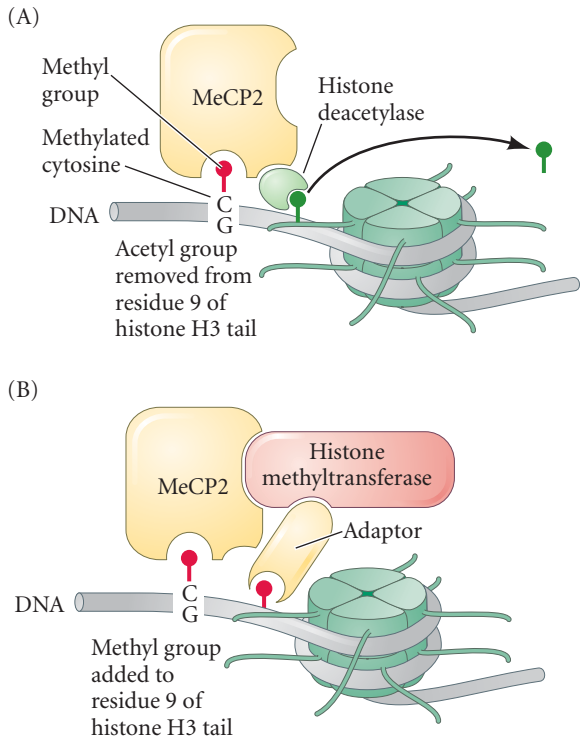


FIGURE 2.19 Modifying nucleosomes through methylated DNA. MeCP2 recognizes the methylated cytosines of DNA. It binds to the DNA and is thereby able to recruit histone deacetylases (which take acetyl groups off the histones) (A) or histone methyltransferases (which add methyl groups to the histones) (B). Both modifications promote the stability of the nucleosome and the tight packing of DNA, thereby repressing gene expression in these regions of DNA methylation. (After Fuks 2003.)

strand of DNA and places methyl groups on the newly synthesized strand opposite it (Figure 2.20; see Bird 2002; Burdge et al. 2007). This is why it is necessary for the C to be next to a G in the sequence. Thus, in each cell division, the pattern of DNA methylation can be maintained. The newly synthesized (unmethylated) strand will become

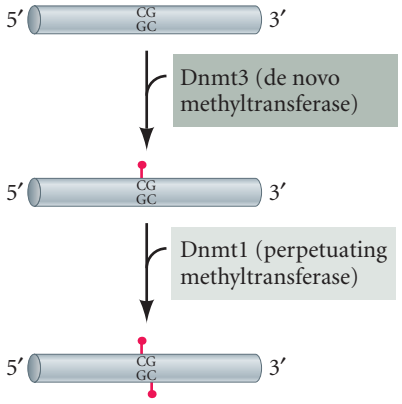


FIGURE 2.20 Two DNA methyltransferases are critically important in modifying DNA. The "de novo" methyltransferase Dnmt3 can place a methyl group on unmethylated cytosines. The "perpetuating" methyltransferase, Dnmt1, recognizes methylated Cs on one strand and methylates the C on the CG pair on the opposite strand.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

properly methylated when Dnmt1 binds to a methylC on the old CpG sequence and methylates the cytosine of the CpG sequence on the complementary strand. In this way, once the DNA methylation pattern is established in a cell, it can be stably inherited by all the progeny of that cell.

Reinforcement between repressive chromatin and repressive DNA has also been observed. Just as methylated DNA is able to attract proteins that deacetylate histones and attract H1 linker histones (both of which will stabilize nucleosomes), so repressive states of chromatin are able to recruit enzymes that methylate DNA. DNA methylation patterns during gametogenesis depend in part on the DNA methyltransferase Dnmt3L. It actually has lost its enzymatic activity, but it can still bind avidly to the amino end of histone H3. However, if the lysine at H3K4 is methylated, it will not bind. Once bound, however, it will recruit and/or activate the DNA methyltransferase Dnmt3A2 to methylate the cytosines on nearby CG pairs (Fan et al. 2007; Ooi et al. 2007).

See WEBSITE 2.7
Silencing large blocks of chromatin

SIDELIGHTS & SPECULATIONS

Consequences of DNA Methylation

The control of transcription through DNA methylation has many consequences in addition to cell differentiation. DNA methylation has explained X chromosome inactivation and DNA imprinting. Moreover, as we will see in Chapter 18, improper DNA methylation (when the wrong cytosines are methylated or demethylated) has been associated with aging, cancers, and the poor health of cloned animals.

X chromosome inactivation

In *Drosophila*, nematodes, and mammals, females are characterized as having two X chromosomes per cell, while males are characterized as having a single X chromosome per cell. Unlike the Y chromosome, the X chromosome contains thousands of genes that are essential for cell activity. Yet despite the female's cells having double the number of X chromosomes, male and female cells contain approximately equal amounts of X chromosome-encoded gene products. This equalization phenomenon is called **dosage compensation**, and it can be accomplished in three ways (Migeon 2002). In *Drosophila*, the transcription rate of the male X chromosomes is doubled so that the single male X chromosome makes the same amount of transcript as the two female X chromosomes (Lucchesi and Manning 1987). This is accomplished by acetylation of the nucleosomes throughout the male's X chromosomes, which gives RNA polymerase more efficient access to that chromosome's promoters (Akhtar et al. 2000; Smith et al. 2001). In *C. elegans*, both X chromosomes are partially repressed (Chu et al. 2002) so that the male and female* products of the X chromosomes are equalized.

In mammals, dosage compensation occurs through the inactivation of one X chromosome in each female cell. Thus, each mammalian somatic cell, whether male or female, has only one functioning X chromosome. This phenomenon is called **X chromosome inactivation**. The chromatin of the inactive X chromosome is converted into **heterochromatin**—chromatin that remains condensed throughout most of the cell cycle and replicates later than most of the other chromatin (the **euchromatin**) of the nucleus. This was first shown by Mary Lyon (1961), who observed coat color patterns in mice. If a mouse is heterozygous for an autosomal gene controlling hair pigmentation, then it resembles one of its two parents, or has a color intermediate between the two. In either case, the mouse is a single color. But if a female mouse is heterozygous for a pigmentation gene *on the X chromosome*, a different result is seen: patches of one parental color alternate with patches of the other parental color. This also explains why calico and tortoiseshell cats† are normally female: their coat color alleles (black and orange) are on the X chromosome (Centerwall and Benirscke 1973).

Lyon proposed the following hypothesis to account for these results:

1. Very early in the development of female mammals, both X chromosomes are active. As development proceeds, one X chromosome is inactivated in each cell (Figure 2.21A).
2. This inactivation is random. In some cells, the paternally derived X chromosome is inactivated; in other cells, the maternally derived X chromosome is shut down.
3. This process is irreversible. Once a particular X chromosome (either the one derived from the mother or the one derived from the father) has been inactivated in a cell, the same X chromosome is inactivated in all of that cell's progeny (Figure 2.21B,C). Because X inactivation happens relatively early in development, an entire region of cells derived from a single cell may all have the same X chromosome inactivated. Thus, all tissues in female mammals are mosaics of two cell types.

*As we will see in Chapter 5, the "female" is actually a hermaphrodite capable of making both sperm and eggs.

†Although the terms *calico* and *tortoiseshell* are sometimes used synonymously, tortoiseshell coats are a patchwork of black and orange only; calico cats usually have white patches—i.e., patches with no pigment—as well (see Figure 2.2).

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

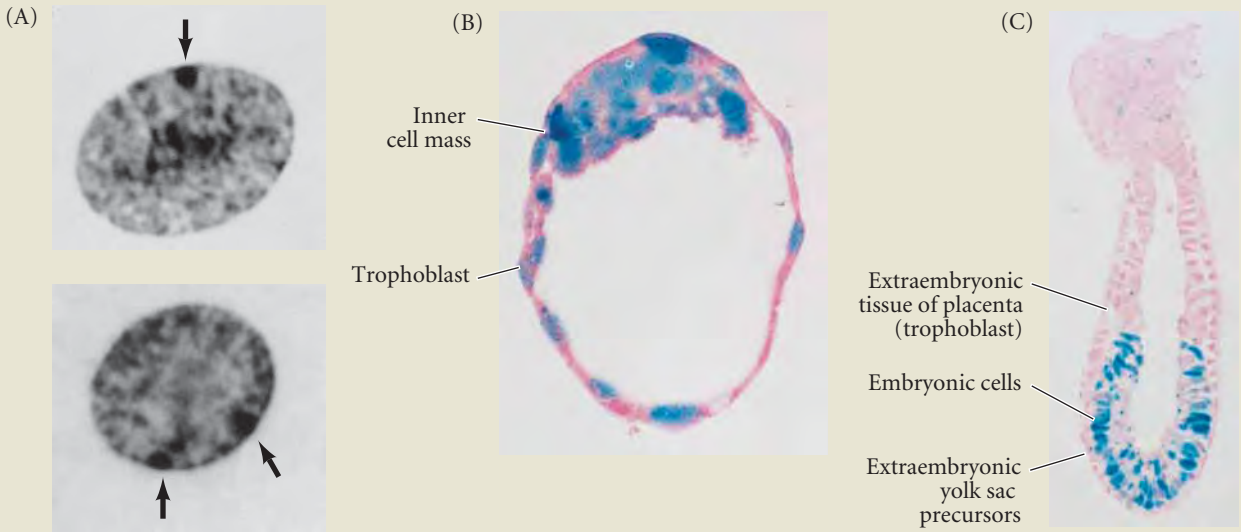


Figure 2.21 X chromosome inactivation in mammals. (A) Inactivated X chromosomes, or *Barr bodies*, in the nuclei of human oral epithelial cells. The top cell is from a normal XX female and has a single Barr body (arrow). In the lower cell, from a female with three X chromosomes, two Barr bodies can be seen. In both cases, only one X chromosome per cell is active. (B,C) The paternally derived X chromosome of this mouse embryo contained a *lacZ* transgene. Those cells in which the chromosome is active make β-galactosidase and stain blue. The other cells are counterstained and appear pink. (B) In the early blastocyst stage (day 4), both X chromosomes are active in all cells. (C) At day 6, random inactivation of one of the chromosomes occurs. Embryonic cells in which the maternal X is active appear pink, while those where the paternal X is active stain blue. In mouse (but not human) trophoblast, the paternally derived X chromosome is preferentially inactivated, so the trophoblast cells are uniformly pink. (A courtesy of M. L. Barr; B,C from Sugimoto et al. 2000, photographs courtesy of N. Takagi.)

The inactivation of the X chromosome is complicated; indeed, it is a bottleneck that many female embryos do not get through (Migeon 2007). The mechanisms of X chromosome inactivation appear to differ between mammalian groups, but these mechanisms converge in that they all inactivate an X chromosome by methylating promoters. In mice and humans, the promoter regions of numerous genes are methylated on the inactive X chromosome and unmethylated on the active X chromosome (Wolf et al. 1984; Keith et al. 1986; Migeon et al. 1991). The memory of this "X inactivation" is transmitted to the progeny of the cells by successive DNA methylation through Dnmt1 (see above).

Genomic Imprinting

The second phenomenon explained by DNA methylation is genomic imprinting. It is usually assumed that the genes one inherits from one's father and the genes one inherits from one's mother are equivalent. In fact, the basis for Mendelian ratios (and the Punnett square analyses used to teach them) is that it does not matter whether the genes came from the sperm or from the egg. But in mammals, there are at least 80 genes for which it *does* matter.* Here, the chromosomes from the male and the female are not equivalent. In these cases, only the sperm-derived or only the egg-derived allele of the gene is expressed. This means that a severe or lethal condition arises if a mutant allele is derived from one parent, but that the same mutant allele will have no deleterious effects if inherited from the other parent. In some of these cases, the nonfunctioning gene has been rendered inactive by DNA methylation. (This means that a mammal must have both a male parent and a female parent. Unlike sea urchins, flies, and frogs, mammals cannot experience parthenogenesis, or "virgin birth.") The methyl groups are placed on the DNA during spermatogenesis and oogenesis by a series of enzymes that first take the existing methyl groups off the chromatin and then place new sex-specific ones on the DNA (Ciccone et al. 2009).

As described earlier in this chapter, methylated DNA is associated with stable DNA silencing, either (1) by interfering with the binding of gene-activating transcription factors or (2) by recruiting repressor proteins that stabilize nucleosomes in a restrictive manner along the gene. The presence of a methyl group in the minor groove of DNA can prevent certain transcription factors from binding to the DNA, thereby preventing the gene from being activated (Watt and Molloy 1988).

For example, during early embryonic development in mice, the *Igf2* gene (for insulin-like growth factor) is active only from the father's chromosome 7. The egg-derived *Igf2* gene does not function during embryonic development. This is because the CTCF protein is an inhibitor that can block the promoter from getting activation signals from enhancers. It binds to a region near the *Igf2* gene in females because this region is not methylated. Once bound, it prevents the maternally derived *Igf2* gene from functioning. In the sperm-derived chromosome 7, the region where CTCF would bind is methylated. CTCF cannot bind and the gene is not inhibited from functioning (**Figure 2.22**; Bartolomei et al. 1993; Ferguson-Smith et al. 1993; Bell and Felsenfeld 2000).

In humans, misregulation of *Igf2* methylation causes Beckwith-

(*Continued on next page*)

*A list of imprinted mouse genes is maintained at http://www.har.mrc.ac.uk/research/genomic_imprinting/introduction.html

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

SIDELIGHTS & SPECULATIONS (Continued)

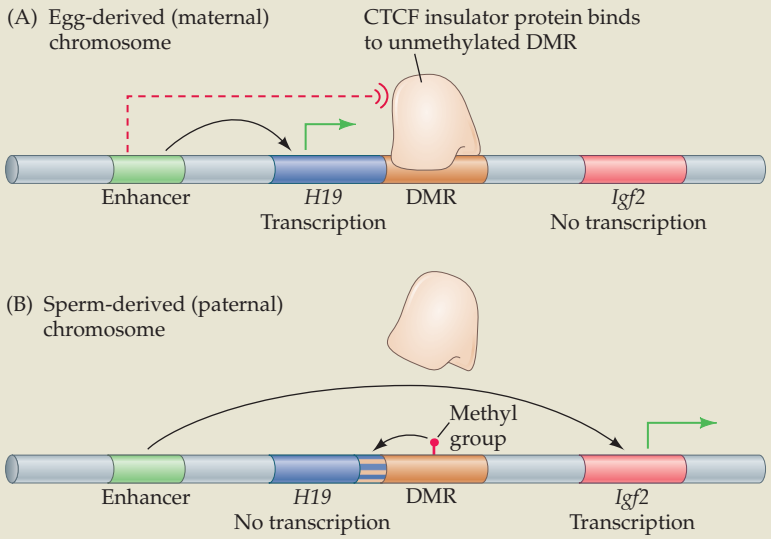


Figure 2.22 Regulation of the imprinted *Igf2* gene in the mouse. This gene is activated by an enhancer element it shares with the *H19* gene. The differentially methylated region (DMR) is a sequence located between the enhancer and the *Igf2* gene, and is found on both sperm- and egg-derived chromosomes. (A) In the egg-derived chromosome, the DMR is unmethylated. The CTCF insulator protein binds to the DMR and blocks the enhancer signal. (B) In the sperm-derived chromosome, the DMR is methylated. The CTCF insulator protein cannot bind to the methylated sequence, and the signal from the enhancer is able to activate *Igf2* transcription.

Wiedemann growth syndrome. Interestingly, although DNA methylation is the mechanism for imprinting this gene in both mice and humans, the mechanisms responsible for the differential *Igf2* methylation between sperm and egg appear to be very different in the two species (Ferguson-Smith et al. 2003; Walter and Paulsen 2003).

Also in humans, the loss of a particular segment of the long arm of chromosome 15 results in different phenotypes, depending on whether the loss is in the male- or the female-derived chromosome (Figure 2.23A). If the chromosome with the defective or missing segment comes from the father, the child is born with Prader-Willi syndrome, a disease associated with mild mental retardation, obesity, small gonads, and short stature. If the defective or missing segment comes from the mother, the child has Angelman syndrome, characterized by severe mental retardation, seizures, lack of speech, and inappropriate laughter (Knoll et al. 1989; Nicholls et al. 1998). The imprinted genes in this region are *SNRPN* and *UBE3A*. In the egg-derived chromosome, *UBE3A* is activated and *SNRPN* is turned off, while in the sperm-derived chromosome, *SNRPN* is activated and *UBE3A* is turned off (Figure 2.23B). The expression of either maternal or paternal loci on human chromosome 15 also depends on methylation differences at specific regions in the chromosome that regulate these genes (Zeschingk et al. 1997; Ferguson-Smith and Surani 2001; Walter and Paulsen 2003).

Differential methylation is one of the most important mechanisms of epigenetic changes. It provides a reminder that an organism cannot be explained solely by its genes. One needs knowledge of developmental parameters as well as genetic ones.

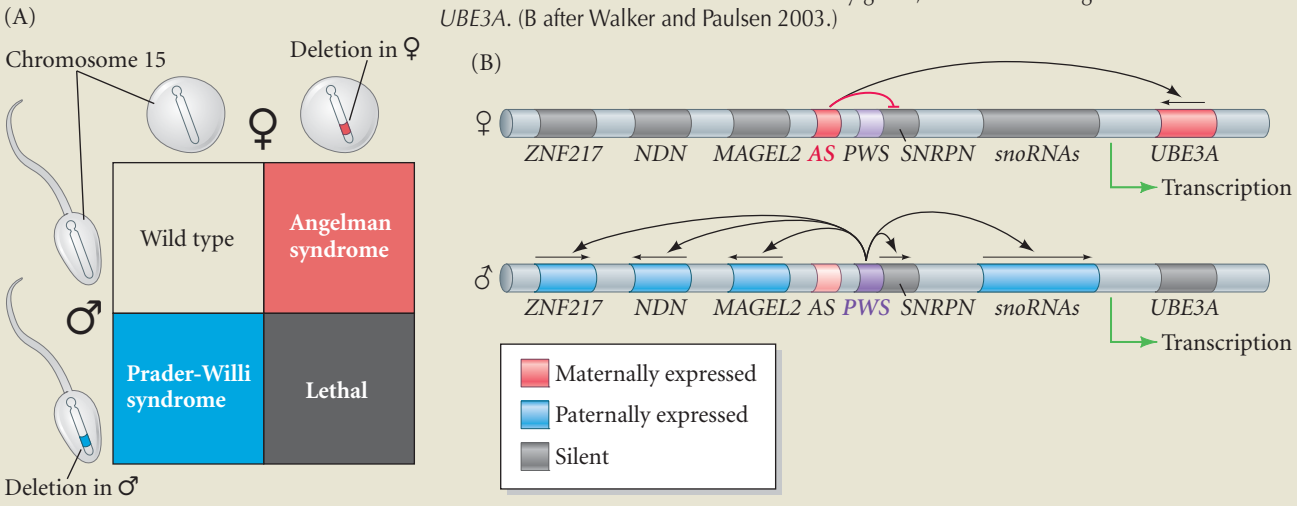


Figure 2.23 Inheritance patterns for Prader-Willi and Angelman syndromes. (A) A region in the long arm of chromosome 15 contains the genes whose absence causes both these syndromes. However, the two conditions are imprinted in reverse fashion. In Prader-Willi syndrome, the paternal genes are active; in Angelman syndrome, the maternal genes are active. (B) Some of the genes and the "inactivation centers" where methylation occurs on this chromosomal region. In the maternal chromosome, the *AS* inactivation center activates *UBE3A* and suppresses *SNRPN*. Conversely, on the paternal chromosome, the *PWS* inactivation center activates *SNRPN* and several other nearby genes, as well as making antisense RNA to *UBE3A*. (B after Walker and Paulsen 2003.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

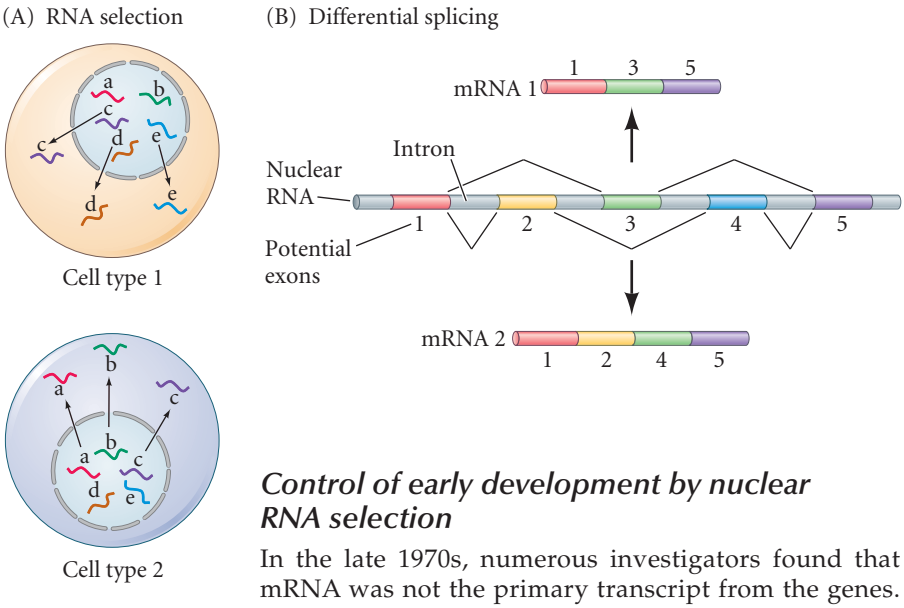


FIGURE 2.24 Roles of differential RNA processing during development. By convention, splicing paths are shown by fine V-shaped lines. (A) RNA selection, whereby the same nuclear RNA transcripts are made in two cell types, but the set that becomes cytoplasmic messenger RNA is different. (B) Differential splicing, whereby the same nuclear RNA is spliced into different mRNAs by selectively using different exons.

Differential RNA Processing

The regulation of gene expression is not confined to the differential transcription of DNA. Even if a particular RNA transcript is synthesized, there is no guarantee that it will create a functional protein in the cell. To become an active protein, the RNA must be (1) processed into a messenger RNA by the removal of introns, (2) translocated from the nucleus to the cytoplasm, and (3) translated by the protein-synthesizing apparatus. In some cases, the synthesized protein is not in its mature form and must be (4) posttranslationally modified to become active. Regulation during development can occur at any of these steps.

The essence of differentiation is the production of different sets of proteins in different types of cells. In bacteria, differential gene expression can be effected at the levels of transcription, translation, and protein modification. In eukaryotes, however, another possible level of regulation exists—namely, control at the level of RNA processing and transport. There are two major ways in which differential RNA processing can regulate development. The first involves "censorship"—selecting which nuclear transcripts are processed into cytoplasmic messages. Different cells select different nuclear transcripts to be processed and sent to the cytoplasm as messenger RNA. Thus, the same pool of nuclear transcripts can give rise to different populations of cytoplasmic mRNAs in different cell types (Figure 2.24A).

The second mode of differential RNA processing is the *splicing* of mRNA precursors into messages that specify different proteins by using different combinations of potential exons. If an mRNA precursor had five potential exons, one cell type might use exons 1, 2, 4, and 5; a different type might use exons 1, 2, and 3; and yet another cell type might use all five (Figure 2.24B). Thus a single gene can produce an entire family of proteins.

Control of early development by nuclear RNA selection

In the late 1970s, numerous investigators found that mRNA was not the primary transcript from the genes. Rather, the initial transcript is a nuclear RNA (nRNA). This nRNA is usually many times longer than the corresponding mRNA because nRNA contains introns that get spliced out during the passage from nucleus to cytoplasm (see Figure 2.6). Originally, investigators thought that whatever RNA was transcribed in the nucleus was processed into cytoplasmic mRNA. But studies of sea urchins showed that different cell types could be *transcribing* the same type of nuclear RNA, but *processing* different subsets of this population into mRNA in different types of cells (Kleene and Humphreys 1977, 1985). Wold and her colleagues (1978) showed that sequences present in sea urchin blastula *messenger* RNA, but absent in gastrula and adult tissue mRNA, were nonetheless present in the *nuclear* RNA of the gastrula and adult tissues.

More genes are transcribed in the nucleus than are allowed to become mRNAs in the cytoplasm. This "censoring" of RNA transcripts has been confirmed by probing for the introns and exons of specific genes. Gagnon and his colleagues (1992) performed such an analysis on the transcripts from the *SpecII* and *CyIIIa* genes of the sea urchin *Strongylocentrotus purpuratus*. These genes encode calcium-binding and actin proteins, respectively, which are expressed only in a particular part of the ectoderm of the sea urchin larva. Using probes that bound to an exon (which is included in the mRNA) and to an intron (which is not included in the mRNA), they found that these genes were being transcribed not only in the ectodermal cells, but also in the mesoderm and endoderm. The analysis of the *CyIIIa* gene showed that the concentration of introns was the same in both the gastrula ectoderm and the mesoderm/endoderm samples, suggesting that this gene was being transcribed at the same rate in the nuclei of all cell types, but was made into cytoplasmic mRNA only in ectodermal cells (Figure 2.25). The unprocessed nRNA for *CyIIIa* is degraded while still in the nuclei of the endodermal and mesodermal cells.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

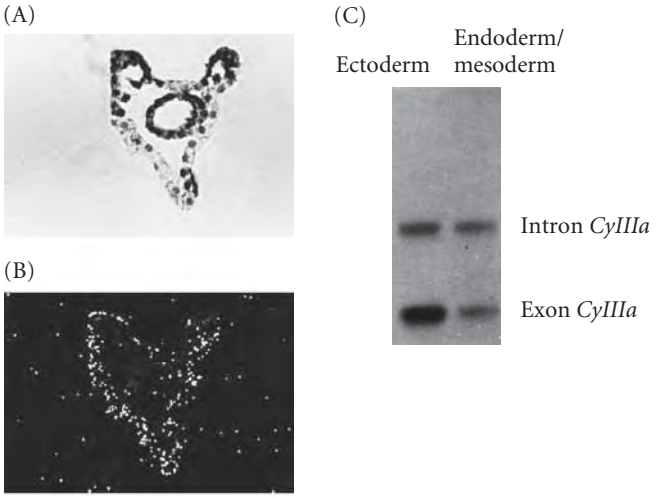


FIGURE 2.25 Regulation of ectoderm-specific gene expression by RNA processing. (A,B) *CyIIIa* mRNA is seen by autoradiography to be present only in the ectoderm. (A) Phase contrast micrograph. (B) In situ hybridization using a probe that binds to a *CyIIIa* exon. (C) The *CyIIIa* nuclear transcript, however, is found in both ectoderm and endoderm/mesoderm. The left lane of the gel represents RNA isolated from the gastrula ectodermal tissue; the right lane represents RNA isolated from endodermal and mesodermal tissues. The upper band is the RNA bound by a probe that binds to an intron sequence (which should be found only in the nucleus) of *CyIIIa*. The lower band represents the RNA bound by a probe complementary to an exon sequence. The presence of the intron indicates that the *CyIIIa* nuclear RNA is being made in both groups of cells, even if the mRNA is seen only in the ectoderm. (From Gagnon et al. 1992; photographs courtesy of R. and L. Angerer.)

Creating families of proteins through differential nRNA splicing

Alternative nRNA splicing is a means of producing a wide variety of proteins from the same gene. The average vertebrate nRNA consists of several relatively short exons (averaging about 140 bases) separated by introns that are usually much longer. Most mammalian nRNAs contain numerous exons. By splicing together different sets of exons, different cells can make different types of mRNAs, and hence, different proteins. Recognizing a sequence of nRNA as either an exon or an intron is a crucial step in gene regulation.

Alternative nRNA splicing is based on the determination of which sequences will be spliced out as introns. This can occur in several ways. Most genes contain "consensus sequences" at the 5′ and 3′ ends of the introns. These sequences are the "splice sites" of the intron. The splicing of nRNA is mediated through complexes known as **spliceosomes** that bind to the splice sites. Spliceosomes are made up of small nuclear RNAs (snRNAs) and proteins called **splicing factors** that bind to splice sites or to the areas adjacent to them. By their production of specific splicing factors, cells can differ in their ability to recognize a sequence as an intron. That is to say, a sequence that is an *exon* in one cell type may be an *intron* in another (Figure 2.26A,B). In other instances, the factors in one cell might recognize different 5′ sites (at the beginning of the intron) or different 3′ sites (at the end of the intron; Figure 2.26C,D).

The 5′ splice site is normally recognized by small nuclear RNA U1 (U1 snRNA) and splicing factor 2 (SF2; also known as alternative splicing factor). The choice of alternative 3′ splice sites is often controlled by which splice site can best bind a protein called U2AF. The spliceosome forms when the proteins that accumulate at the 5′ splice site contact those proteins bound to the 3′ splice site. Once the 5′ and 3′ ends are brought together, the intervening intron is excised and the two exons are ligated together.

Researchers estimate that approximately 92% of human genes are alternatively spliced, and that such alternative splicing is a major way by which the rather limited number of genes can create a much larger array of proteins (Wang et al. 2008). The deletion of certain potential exons in some cells but not in others enables one gene to create a family of closely related proteins. Instead of one gene-one polypeptide, one can have one gene-one family of proteins. For instance, alternative RNA splicing enables the gene for α-tropomyosin to encode brain, liver, skeletal muscle, smooth muscle, and fibroblast forms of this protein (Breitbart et al. 1987). The nuclear RNA for α-tropomyosin contains 11 potential exons, but different sets of exons are used in different cells (Figure 2.27). Such different proteins encoded by the same gene are called **splicing isoforms** of the protein.

In some instances, alternatively spliced RNAs yield proteins that play similar yet distinguishable roles in the same cell. Different isoforms of the WT1 protein perform different functions in the development of the gonads and kidneys. The isoform without the extra exon functions as a transcription factor during kidney development, whereas the isoform containing the extra exon appears to be involved in splicing different nRNAs and may be critical in testis development (Hammes et al. 2001; Hastie 2001).

The *Bcl-x* gene provides a good example of how alternative nRNA splicing can make a huge difference in a protein's function. If a particular DNA sequence is used as an exon, the "large Bcl-X protein," or Bcl-X_L, is made (see Figure 2.26C). This protein inhibits programmed cell death. However, if this sequence is seen as an intron, the "small Bcl-X protein" (Bcl-X_S) is made, and this protein *induces* cell death. Many tumors have a higher than normal amount of Bcl-X_L.

If you get the impression from this discussion that a gene with dozens of introns could create literally thousands of different, related proteins through differential splicing, you are probably correct. The current champion at making multiple proteins from the same gene is the *Drosophila Dscam1* gene. This gene encodes a membrane receptor protein involved in preventing dendrites from the same neuron

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

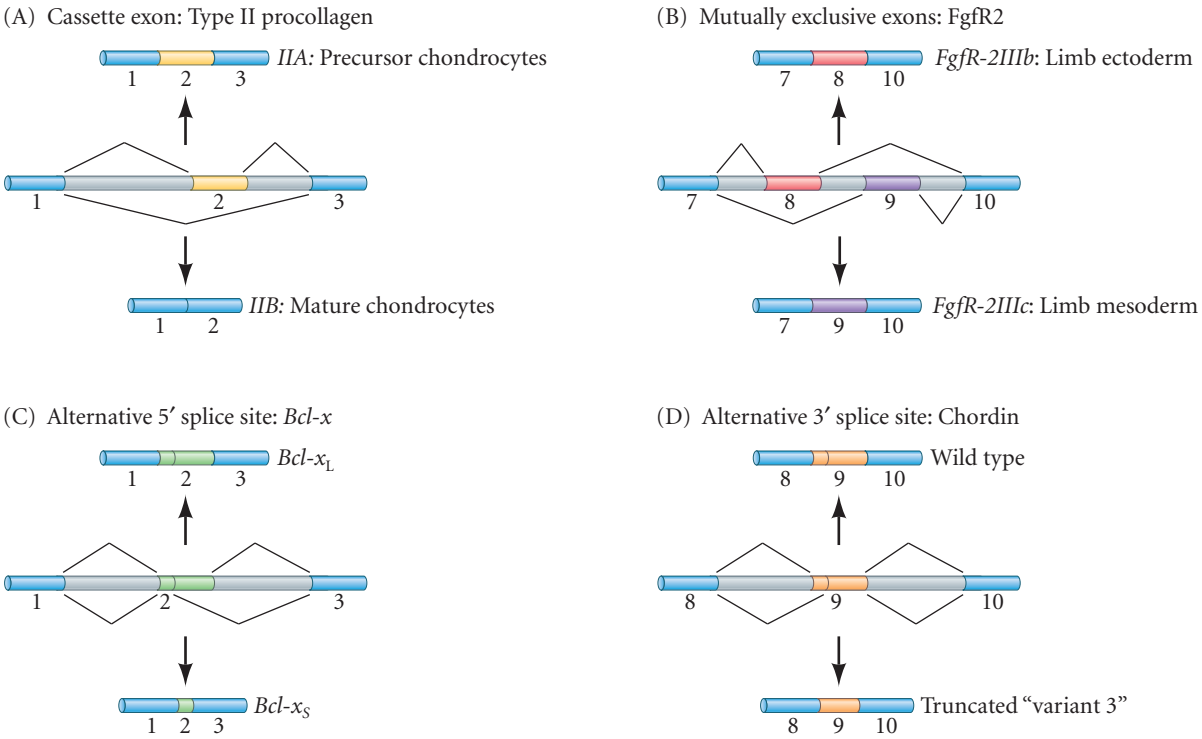


FIGURE 2.26 Some examples of alternative RNA splicing. Blue and colored portions of the bars represent exons; gray represents introns. Alternative splicing patterns are shown with V-shaped lines. (A) A "cassette" (yellow) that can be used as exon or removed as an intron distinguishes the type II collagen types of chondrocyte precursors and mature chondrocytes (cartilage cells). (B) Mutually exclusive exons distinguish fibroblast growth factor receptors found in the limb ectoderm from those found in the limb mesoderm. (C) Alternative 5′ splice site selection, such as that used to create the large and small isoforms of the protein Bcl-X. (D) Alternative 3′ splice sites are used to form the normal and truncated forms of chordin. (After McAlinden et al. 2004.)

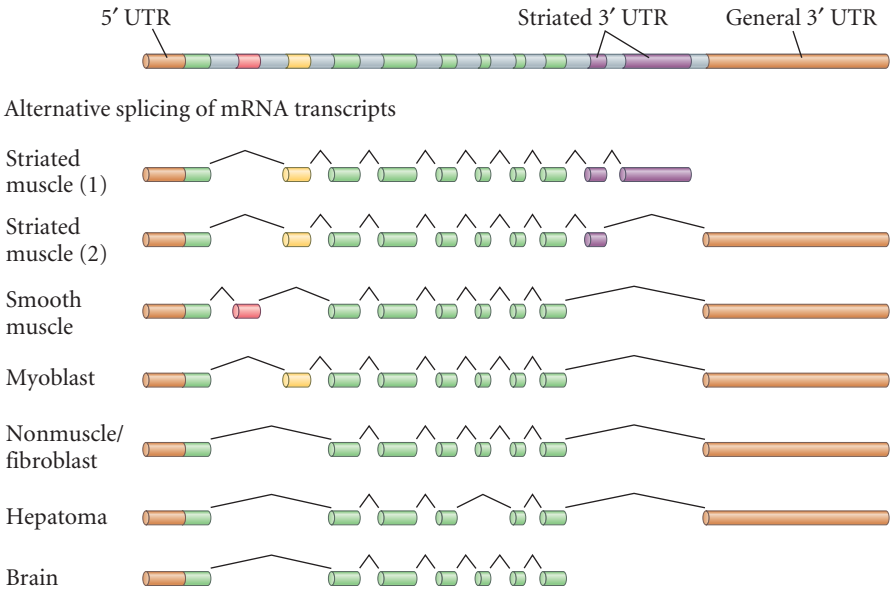


FIGURE 2.27 Alternative RNA splicing to form a family of rat α-tropomyosin proteins. The α-tropomyosin gene is represented on top. The thin lines represent the sequences that become introns and are spliced out to form the mature mRNAs. Constitutive exons (found in all tropomyosins) are shown in green. Those expressed only in smooth muscle are red; those expressed only in striated muscle are purple. Those that are variously expressed are yellow. Note that in addition to the many possible combinations of exons, two different 3′ ends ("striated" and "general") are possible. (After Breitbart et al. 1987.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

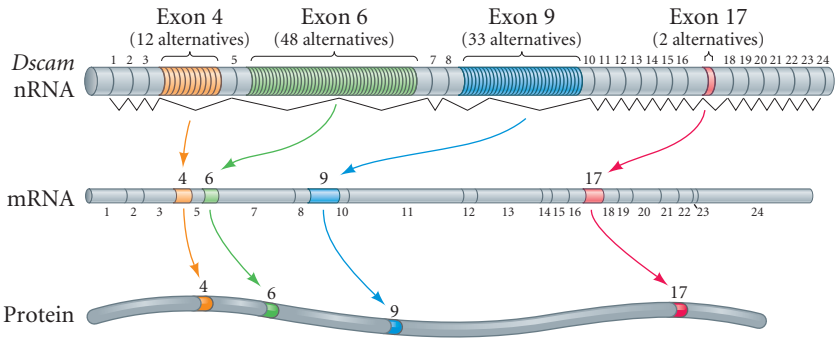


FIGURE 2.28 The *Dscam* gene of *Drosophila* can produce 38,016 different types of proteins by alternative nRNA splicing. The gene contains 24 exons. Exons 4, 6, 9, and 17 are encoded by sets of mutually exclusive possible sequences. Each messenger RNA will contain one of the 12 possible exon 4 sequences, one of the 48 possible exon 6 alternatives, one of the 33 possible exon 9 alternatives, and one of the 2 possible exon 17 sequences. The *Drosophila Dscam* gene is homologous to a DNA sequence on human chromosome 21 that is expressed in the nervous system. Disturbances of this gene in humans may lead to the neurological defects of Down syndrome (Yamakawa et al. 1998; Saito 2000).

from binding to one another. *Dscam1* contains 115 exons. However, a dozen different adjacent DNA sequences can be selected to be exon 4. Similarly, more than 30 mutually exclusive adjacent DNA sequences can become exons 6 and 9, respectively (Figure 2.28; Schmucker et al. 2000). If all possible combinations of exons are used, this one gene can produce 38,016 different proteins, and random searches for these combinations indicate that a large fraction of them are in fact made. The nRNA of *Dscam1* has been found to be alternatively spliced in different axons, and when two dendrites from the same axon touch each other, they are repelled. This causes the extensive branching of the dendrites. It appears that the thousands of splicing isoforms are needed to ensure that each neuron acquires a unique identity (Figure 2.29; Schmucker 2007; Millard and Zipursky 2008; Hattori et al. 2009). The *Drosophila* genome is thought to contain only 14,000 genes, but here is a single gene that encodes three times that number of proteins!

About 92% of human genes are thought to produce multiple types of mRNA. Therefore, even though the human genome may contain 20,000–30,000 genes, its **proteome**—the number and type of proteins encoded by the genome—is far more complex. "Human genes are multitaskers," notes Christopher Burge, one of the scientists who calculated this figure (Ledford 2008). This explains an important paradox. *Homo sapiens* has around 20,000 genes in each nucleus; so does the nematode *Caenorhabditis elegans*, a tubular creature with only 969 cells. We have more cells and cell types in the shaft of a hair than *C. elegans* has in its entire body. What's this worm doing with the same number of genes as us? The answer is that *C. elegans* genes rarely make isoforms. Each gene in the worm makes but one protein, whereas in humans the same number of genes produces an enormous array of different proteins.

Splicing enhancers and recognition factors

The mechanisms of differential RNA processing involve both *cis*-acting sequences on the nRNA and *trans*-acting protein factors that bind to these regions (Black 2003). The *cis*-acting sequences on nRNA are usually close to their potential 5′ or 3′ splice sites. These sequences are called "splicing enhancers," since they promote the assembly of spliceosomes at RNA cleavage sites. Conversely, these same sequences can be "splicing silencers" if they act to exclude exons from an mRNA sequence. These sequences are recognized by *trans*-acting proteins, most of which can recruit spliceosomes to that area. However, some *trans*-acting proteins, such as the polypyrimidine tract-binding protein (PTP),* repress spliceosome formation where they bind.

As might be expected, there are some splicing enhancers that appear to be specific for certain tissues. Muscle-specific *cis*-regulatory sequences have been found around those exons characterizing muscle cell messages. These are recognized by certain proteins that are found in the muscle cells early in their development (Ryan and Cooper 1996; Charlet-B et al. 2002). Their presence is able to compete with the PTP that would otherwise prevent the inclusion of the muscle-specific exon into the mature message. In this way, an entire battery of muscle-specific isoforms can be generated.

*PTP is involved in making the correct isoform of tropomyosin and may be especially important in determining the mRNA populations of the brain. PTP is also involved in the mutually exclusive use of exon IIIb or IIIc in the mRNA for fibroblast growth factor 2 (see Figure 2.26B; Carstens et al. 2000; Lilleväli et al. 2001; Robinson and Smith 2006).

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

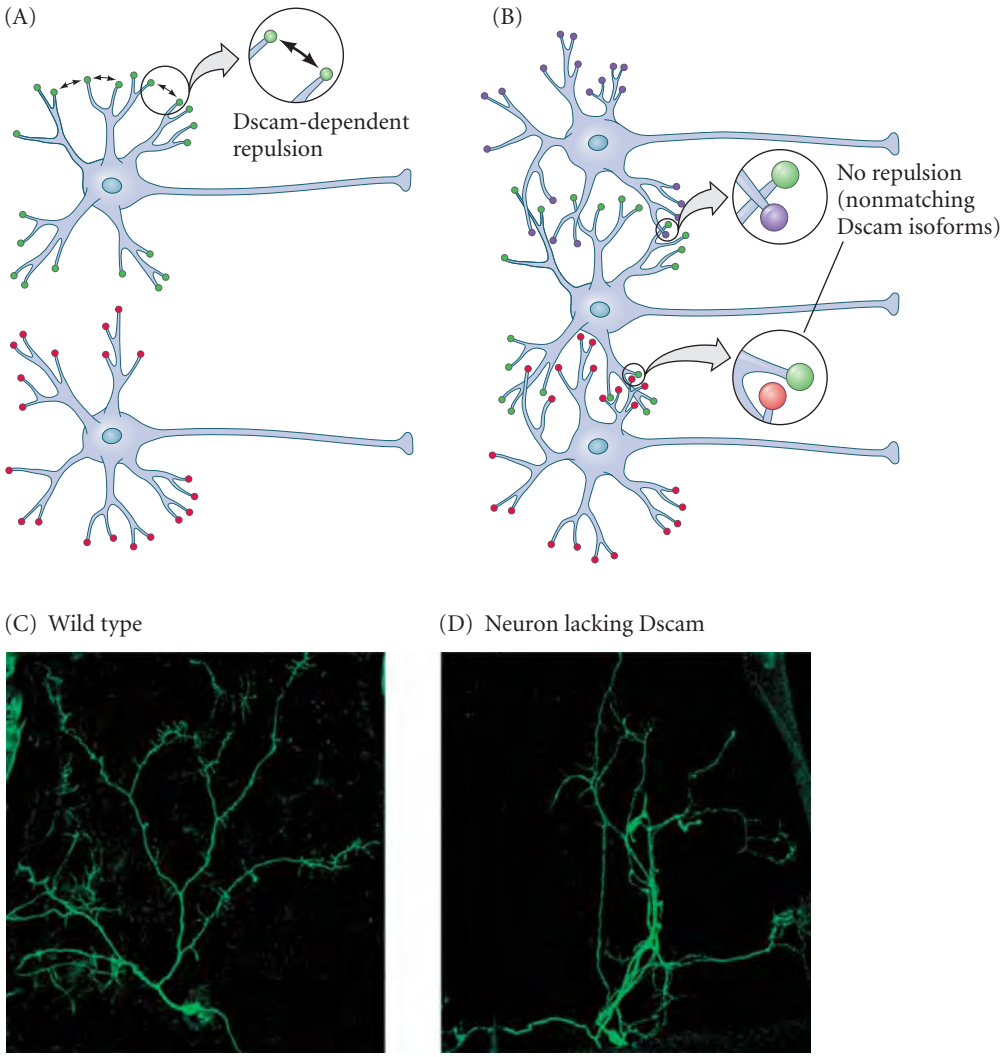


FIGURE 2.29 Dscam protein is specifically required to keep dendrites from the same neuron from adhering to each other. (A) When sister dendrites expressing the same splicing isoform of Dscam touch, the Dscam-Dscam interactions repel them and cause the dendrites to separate. (B) Different neurons express different splicing isoforms that do not interact with one another (and therefore do not trigger repulsion), which allows neurons to interact. (C) Neurons with multiple dendrites normally develop highly branched formations in which none of the branches crosses another. (D) Loss of *Dscam1* in such a neuron abolishes self-repulsion and results in excessive self-crossing and adhesion. (After Schmucker 2007; photographs courtesy of Dietmar Schmucker.)

One might also suspect that mutations of the splicing sites would lead to alternative phenotypes. Most of these splice site mutations lead to nonfunctional proteins and serious diseases. For instance, a single base change at the 5′ end of intron 2 in the human β-globin gene prevents splicing from occurring and generates a nonfunctional mRNA (Baird et al. 1981). This causes the absence of any β-globin from this gene, and thus a severe (and often life-threatening) type of anemia. Similarly, a mutation in the *dystrophin* gene at a particular splice site causes the skipping of that exon and a severe form of muscular dystrophy (Sironi et al. 2001). In at least one such case, the splice site mutation was not dangerous and actually gave the patient greater strength. In a different case, Schuelke and colleagues (2004) described a family in which individuals in four generations had a splice site mutation in the *myostatin* gene (Figure 2.30A). Among the family members were professional athletes and a 4-year-old toddler who was able to hold two 3-kg dumb-

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

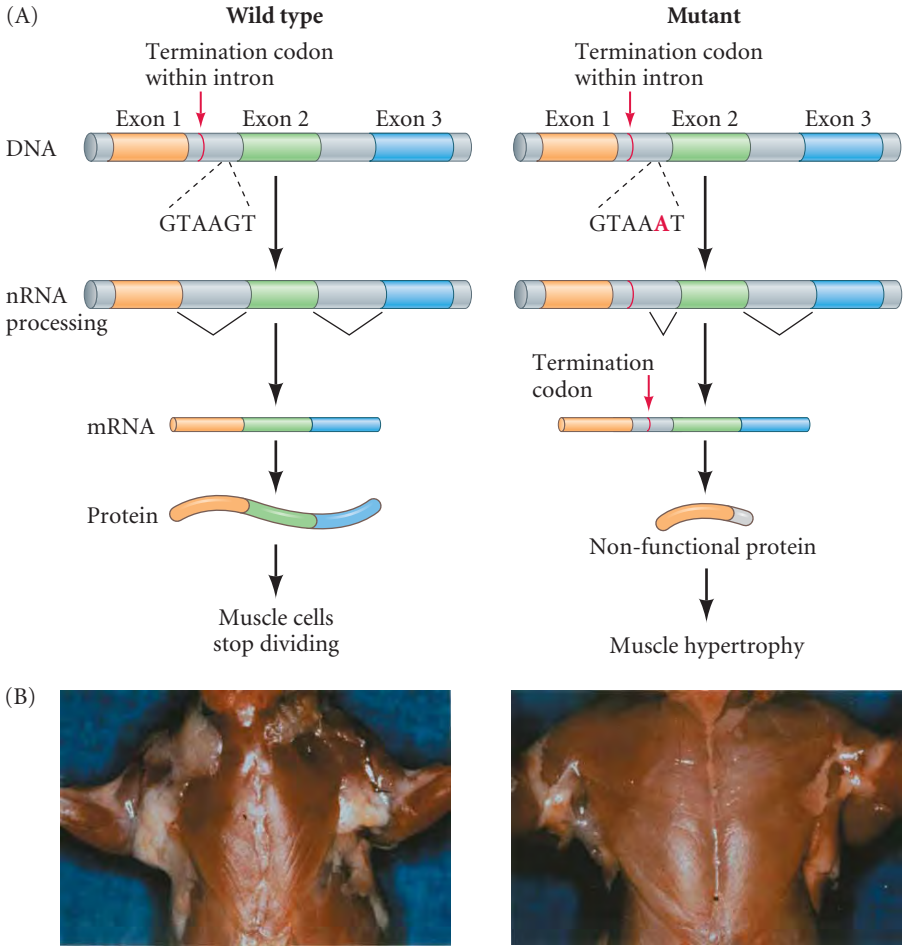


FIGURE 2.30 Muscle hypertrophy through mispliced RNA. This mutation results in a deficiency of the negative growth regulator myostatin in the muscle cells. (A) Molecular analysis of the mutation. There is no mutation in the coding sequence of the gene, but in the first intron, a mutation from a G to an A created a new (and widely used) splicing site. This caused aberrant nRNA splicing and the inclusion of an early protein synthesis termination codon into the mRNA. Thus, proteins made from that message would have been short and nonfunctional. (B) Pectoral musculature of a "mighty mouse" with the mutation (right) compared to the muscles of a wild-type mouse (left). (A after Schuelke et al. 2004; B from McPherron et al. 1997; courtesy of A. C. McPherron.)

bells with his arms fully extended. The *myostatin* gene product is a negative regulator—a factor that tells muscle precursor cells to stop dividing. In mammals with the mutation, the muscles are not told to differentiate until they have undergone many more rounds of cell division, and the result is larger muscles (Figure 2.30B).

Control of Gene Expression at the Level of Translation

The splicing of nuclear RNA is intimately connected with its export through the nuclear pores and into the cytoplasm. As the introns are removed, specific proteins bind to the spliceosome and attach the spliceosome-RNA complex to nuclear pores (Luo et al. 2001; Strässer and Hurt 2001). But once the RNA has reached the cytoplasm, there is still no guarantee that it will be translated. The control of gene expression at the level of translation can occur by many means; some of the most important of these are described below.

Differential mRNA longevity

The longer an mRNA persists, the more protein can be translated from it. If a message with a relatively short half-life were selectively stabilized in certain cells at certain times, it would make large amounts of its particular protein only at those times and places.

The stability of a message often depends on the length of its polyA tail. This, in turn, depends largely on sequences in the 3′ untranslated region, certain of which allow longer polyA tails than others. If these 3′ UTR regions are experimentally traded, the half-lives of the resulting mRNAs are altered: long-lived messages will decay rapidly, while normally short-lived mRNAs will remain around longer (Shaw and Kamen 1986; Wilson and Treisman 1988; Decker and Parker 1995).

In some instances, messenger RNAs are selectively stabilized at specific times in specific cells. The mRNA for casein, the major protein of milk, has a half-life of 1.1 hours in rat mammary gland tissue. However, during periods of lactation, the presence of the hormone prolactin increases this half-life to 28.5 hours (Figure 2.31; Guyette et al. 1979). In the development of the nervous system, a group of proteins called HuD proteins stabilizes a group of mRNAs that stop the neuronal precursor cells from dividing and also stabilizes a second group of mRNAs that are critical for these cells to start neuron differentiation (Okano and Darnell 1997; Descchenes-Furry et al. 2006, 2007).

Selective inhibition of mRNA translation: Stored oocyte mRNAs

Some of the most remarkable cases of translational regulation of gene expression occur in the oocyte. The oocyte often makes and stores mRNAs that will be used only after

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

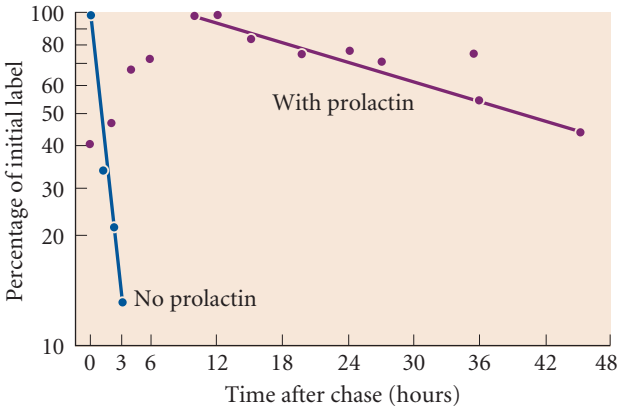


FIGURE 2.31 Degradation of casein mRNA in the presence and absence of prolactin. Cultured rat mammary cells were given radioactive RNA precursors (pulse) and, after a given time, were washed and given nonradioactive precursors (chase). This procedure labeled the casein mRNA synthesized during the pulse time. Casein mRNA was then isolated at different times following the chase and its radioactive label measured. In the absence of prolactin, the labeled (i.e., newly synthesized) casein mRNA decayed rapidly, with a half-life of 1.1 hours. When the same experiment was done in a medium containing prolactin, the half-life was extended to 28.5 hours. (After Guyette et al. 1979.)

fertilization occurs. These messages stay in a dormant state until they are activated by ion signals (discussed in Chapter 4) that spread through the egg during ovulation or fertilization.

Table 2.2 gives a partial list of mRNAs that are stored in the oocyte cytoplasm. Some of these stored mRNAs encode proteins that will be needed during cleavage, when the embryo makes enormous amounts of chromatin, cell membranes, and cytoskeletal components. Some of them encode cyclin proteins that regulate the timing of early cell division (Rosenthal et al. 1980; Standart et al. 1986). Indeed, in many species (including sea urchins and *Drosophila*),

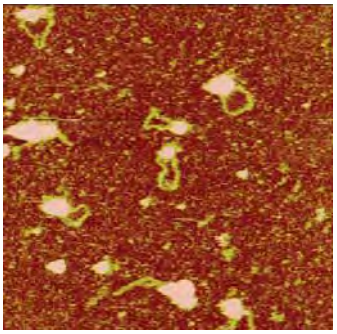
TABLE 2.2 Some mRNAs stored in oocyte cytoplasm and translated at or near fertilization

mRNAs encoding	Function(s)	Organism(s)
Cyclins	Cell division regulation	Sea urchin, clam, starfish, frog
Actin	Cell movement and contraction	Mouse, starfish
Tubulin	Formation of mitotic spindles, cilia, flagella	Clam, mouse
Small subunit of ribonucleotide reductase	DNA synthesis	Sea urchin, clam, starfish
Hypoxanthine phosphoribosyl-transferase	Purine synthesis	Mouse
Vg1	Mesodermal determination(?)	Frog
Histones	Chromatin formation	Sea urchin, frog, clam
Cadherins	Blastomere adhesion	Frog
Metalloproteinases	Implantation in uterus	Mouse
Growth factors	Cell growth; uterine cell growth(?)	Mouse
Sex determination factor FEM-3	Sperm formation	*C. elegans*
PAR gene products	Segregate morphogenetic determinants	*C. elegans*
SKN-1 morphogen	Blastomere fate determination	*C. elegans*
Hunchback morphogen	Anterior fate determination	*Drosophila*
Caudal morphogen	Posterior fate determination	*Drosophila*
Bicoid morphogen	Anterior fate determination	*Drosophila*
Nanos morphogen	Posterior fate determination	*Drosophila*
GLP-1 morphogen	Anterior fate determination	*C. elegans*
Germ cell-less protein	Germ cell determination	*Drosophila*
Oskar protein	Germ cell localization	*Drosophila*
Ornithine transcarbamylase	Urea cycle	Frog
Elongation factor 1a	Protein synthesis	Frog
Ribosomal proteins	Protein synthesis	Frog, *Drosophila*

Compiled from numerous sources

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

(A) Circularized mRNA



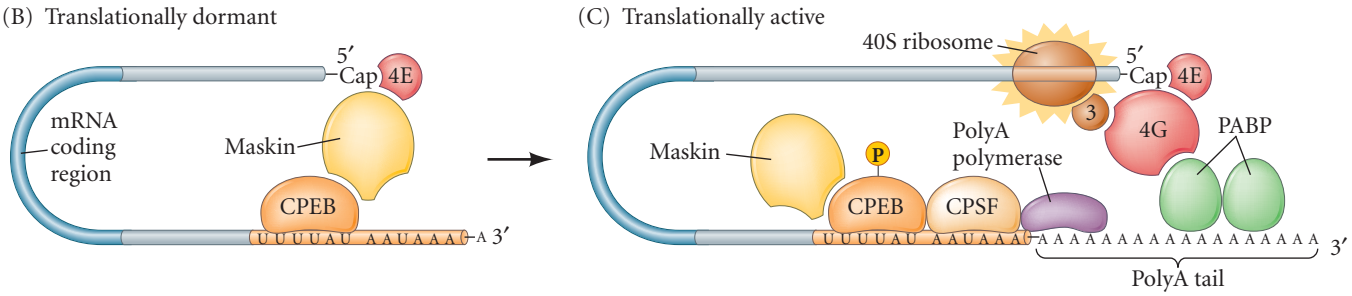


FIGURE 2.32 Translational regulation in oocytes. (A) Messenger RNAs are often found as circles, where the 5′ end and the 3′ end contact one another. Here, a yeast mRNA seen by atomic force microscopy is circularized by eIF4E and eIF4G (5′ end) and the polyA binding protein (3′ end). (B) In *Xenopus* oocytes, the 3′ and 5′ ends of the mRNA are brought together by maskin, a protein that binds to CPEB on the 3′ end and translation initiation factor 4E (eIF4E) on the 5′ end. Maskin blocks the initiation of translation by preventing eIF4E from binding eIF4G. (C) When stimulated by progesterone during ovulation, a kinase phosphorylates CPEB, which can then bind CPSF. CPSF can bind polyA polymerase and initiate growth of the polyA tail. PolyA binding protein (PABP) can bind to this tail and then bind eIF4G in a stable manner. This initiation factor can then bind eIF4E and, through its association with eIF3, position a 40S ribosomal subunit on the mRNA. (A from Wells et al. 1998; B,C after Mendez and Richter 2001.)

maintenance of the normal rate and pattern of early cell divisions does not require a nucleus; rather, it requires continued protein synthesis from stored maternal mRNAs (Wagenaar and Mazia 1978; Edgar et al. 1994). Other stored messages encode proteins that determine the fates of cells. These include the *bicoid*, *caudal*, and *nanos* messages that provide information in the *Drosophila* embryo for the production of its head, thorax, and abdomen.

Most translational regulation in oocytes is negative, as the "default state" of the mRNA is to be available for translation. Therefore, there must be inhibitors preventing the translation of these mRNAs in the oocyte, and these inhibitors must somehow be removed at the appropriate times around fertilization. The 5′ cap and the 3′ untranslated region seem especially important in regulating the accessibility of mRNA to ribosomes. If the 5′ cap is not made or if the 3′ UTR lacks a polyA tail, the message probably will not be translated. The oocytes of many species have "used these ends as means" to regulate the translation of their mRNAs.

It is important to realize that, unlike the usual representations of mRNA, most mRNAs probably form circles, with their 3′ end being brought to their 5′ end (Figure 2.32A). The 5′ cap is bound by **eukaryotic initiation factor-4E** (**eIF4E**), a protein that is also bound to eIF4A (a helicase that unwinds double-stranded regions of RNA) and **eIF4G**, a scaffold protein that allows the mRNA to bind to the ribosome through its interaction with eIF4E (Wells et al. 1998; Gross et al. 2003). The polyA binding protein, which sits on the polyA tail of the mRNA, also binds to the eIF4G protein. This brings the 3′ end of the message next to the 5′ end and allows the messenger RNA to be recognized by the ribosome. Thus, the 5′ cap is critical for translation, and some animal's oocytes have used this as a direct means of translational control. For instance, the oocyte of the tobacco hornworm moth makes some of its mRNAs without their methylated 5′ caps. In this state, they cannot be efficiently translated. However, at fertilization, a methyltransferase completes the formation of the caps, and these mRNAs can be translated (Kastern et al. 1982).

In amphibian oocytes, the 5′ and 3′ ends of many mRNAs are brought together by a protein called **maskin** (Stebbins-Boaz et al. 1999; Mendez and Richter 2001). Maskin links the 5′ and 3′ ends into a circle by binding to two other proteins, each at opposite ends of the message. First, it binds to the **cytoplasmic polyadenylation-element-binding protein** (**CPEB**) attached to the UUUUAU sequence in the 3′ UTR; second, maskin also binds to the eIF4E factor that is attached to the cap sequence. In this configuration, the mRNA cannot be translated (Figure 2.32B). The binding of eIF4E to maskin is thought to prevent the binding of eIF4E to eIF4G, a critically important translation initiation factor that brings the small ribosomal subunit to the mRNA.

Mendez and Richter (2001) have proposed an intricate scenario to explain how mRNAs bound together by maskin become translated at about the time of fertilization. At ovulation (when the hormone progesterone stimulates the last meiotic divisions of the oocyte and the oocyte is released for fertilization), a kinase activated by progesterone phosphorylates the CPEB protein. The phosphorylated CPEB can now bind to CPSF, the cleavage and polyadenylation specificity factor (Mendez et al. 2000; Hodgman et al. 2001). The bound CPSF protein sits on a

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

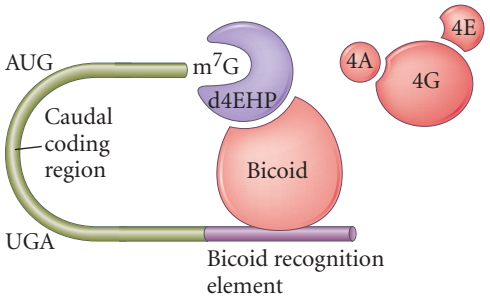


FIGURE 2.33 Protein binding in *Drosophila* oocytes. Bicoid protein binds to a recognition element in the 3′ UTR of the *caudal* message. Bicoid can bind to d4EHP, which prevents the binding of eIF4E to the cap structure. Without eIF4E, the eIF4G cannot bind and initiate translation. (After Cho et al. 2005.)

particular sequence of the 3′ UTR that has been shown to be critical for polyadenylation, and it complexes with a polymerase that elongates the polyA tail of the mRNA. In oocytes, a message having a short polyA tail is not degraded; however, such messages are not translated.

Once the tail is extended, molecules of the polyA binding protein (PABP) can attach to the growing tail. PABP proteins stabilize eIF4G, allowing it to outcompete maskin for the binding site on the eIF4E protein at the 5′ end of the mRNA. The eIF4G protein can then bind eIF3, which can position the small ribosomal subunit onto the mRNA. The small (40S) ribosomal subunit will then find the initiator tRNA, complex with the large ribosomal subunit, and initiate translation (Figure 2.32C).

In the *Drosophila* oocyte, Bicoid can act both as a transcription factor (activating genes such as *hunchback*) and also as a translational inhibitor (see Chapter 6). Bicoid represses the translation of *caudal* mRNA, preventing its transcription in the anterior half of the embryo. (The protein made from the *caudal* message is important in activating those genes that specify the cells to be abdomen precursors.) Bicoid inhibits *caudal* mRNA translation by binding to a "bicoid recognition element," a series of nucleotides in the 3′ UTR of the *caudal* message (Figure 2.33). Once there, Bicoid can bind with and recruit another protein, d4EHP. The d4EHP protein can compete with eIF4E for the cap. Without eIF4E, there is no association with eIF4G and the *caudal* mRNA becomes untranslatable. As a result, the *caudal* message is not translated in the anterior of the embryo (where Bicoid is abundant), but is active in the posterior portion of the embryo.

microRNAs: Specific regulators of mRNA translation and transcription

If proteins can bind to specific nucleic acid sequences to block transcription or translation, you would think that RNA would do the job even better. After all, RNA can be made specifically to bind a particular sequence. Indeed, one of the most efficient means of regulating the translation of a specific message is to make a small RNA complementary to a portion of a particular mRNA. Such a naturally occurring antisense RNA was first seen in *C. elegans* (Lee et al. 1993; Wightman et al. 1993). Here, the *lin-4* gene was found to encode a 21-nucleotide RNA that bound to multiple sites in the 3′ UTR of the *lin-14* mRNA (Figure 2.34). The *lin-14* gene encodes a transcription factor, LIN-14, that is important during the first larval phase of *C. elegans* development. It is not needed afterward, and *C. elegans* is able to inhibit synthesis of LIN-14 from these messages by activating the *lin-4* gene. The binding of *lin-*

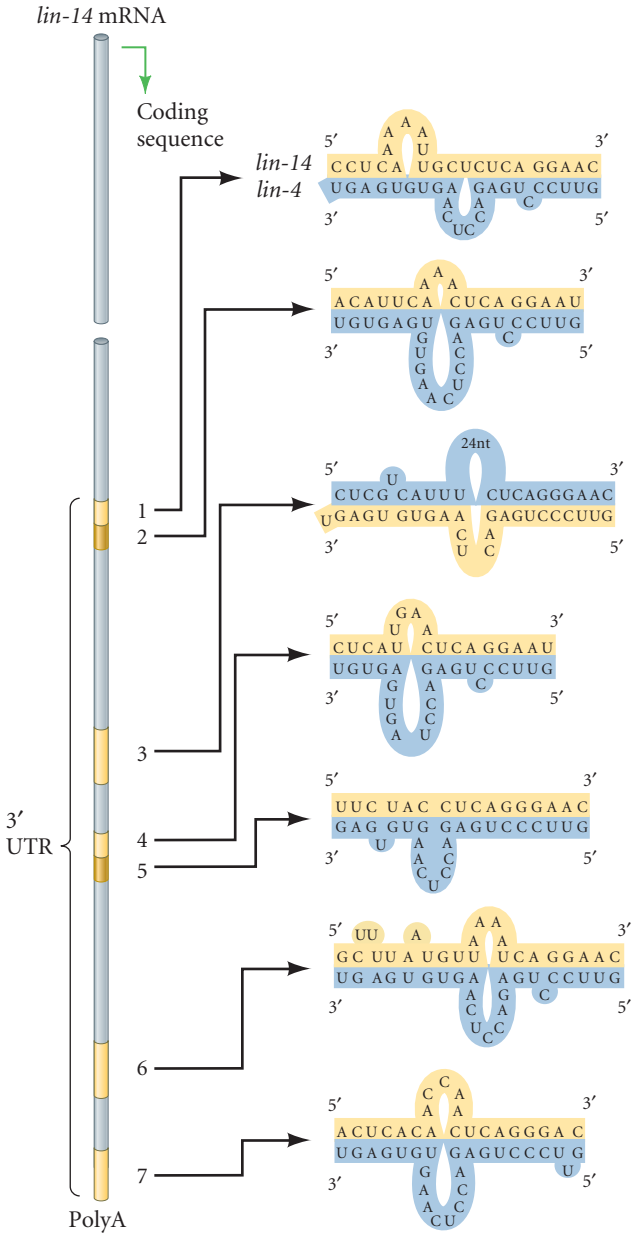


FIGURE 2.34 Hypothetical model of the regulation of *lin-14* mRNA translation by *lin-4* RNAs. The *lin-4* gene does not produce an mRNA. Rather, it produces small RNAs that are complementary to a repeated sequence in the 3′ UTR of the *lin-14* mRNA, which bind to it and prevent its translation. (After Wickens and Takayama 1995.)

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

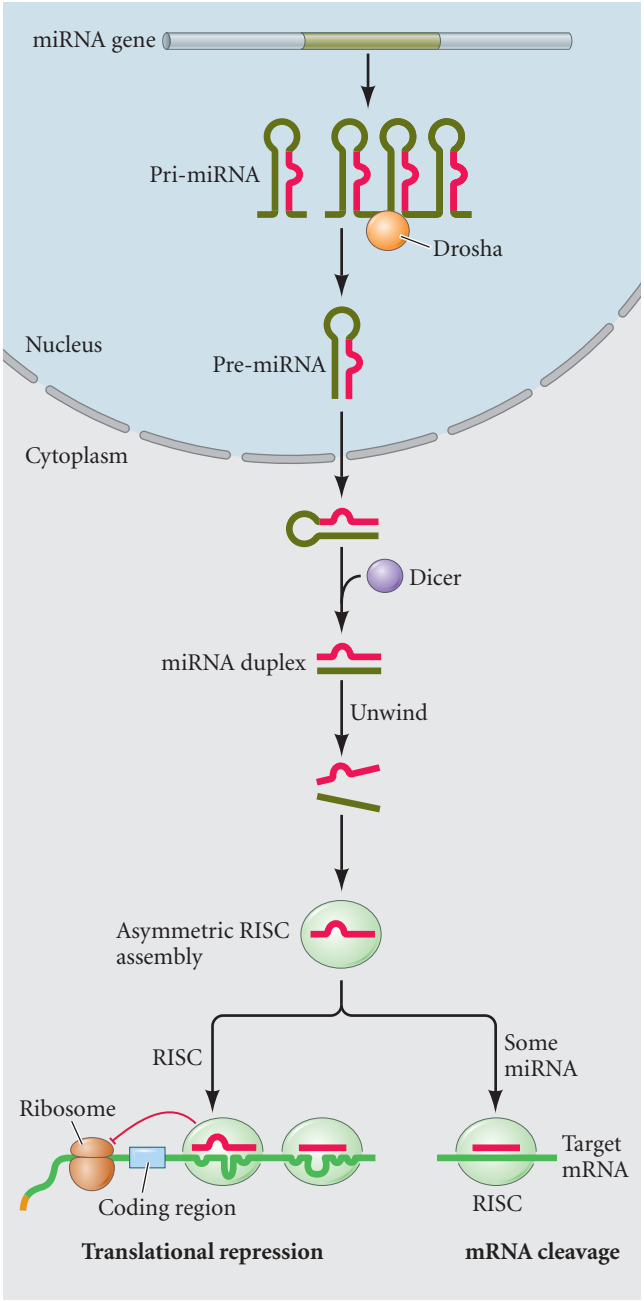


FIGURE 2.35 Current model for the formation and use of microRNAs. The miRNA gene encodes a pri-miRNA that often has several hairpin regions where the RNA finds nearby complementary bases with which to pair. The pri-miRNA is processed into individual pre-miRNA "hairpins" by the Drosha RNAase, and these are exported from the nucleus. Once in the cytoplasm, another RNAase, Dicer, eliminates the non-base-paired loop. Dicer also acts as a helicase to separate the strands of the double-stranded miRNA. One strand (probably recognized by placement of Dicer) is packaged with proteins into the RNA-induced silencing complex (RISC), which subsequently binds to the 3′ UTRs to effect translational suppression or cleavage, depending (at least in part) on the strength of the complementarity between the miRNA and its target. (After He and Hannon 2004.)

4 transcripts to the *lin-14* mRNA 3′ UTR causes degradation of the *lin-14* message (Bagga et al. 2005).

The *lin-4* RNA is now thought to be the "founding member" of a very large group of **microRNAs** (**miRNAs**). These RNAs of about 22 nucleotides are made from longer precursors. These precursors can be in independent transcription units (the *lin-4* gene is far apart from the *lin-14* gene), or they can reside in the introns of other genes (Aravin et al. 2003; Lagos-Quintana et al. 2003). Many of the newly discovered microRNAs have been found in the regions between genes (regions previously considered to contain "junk DNA"). The initial RNA transcript (which may contain several repeats of the miRNA sequence) forms hairpin loops wherein the RNA finds complementary structures within its strand. These stem-loop structures are processed by a set of RNases (Drosha and Dicer) to make single-stranded microRNA (Figure 2.35). The microRNA is then packaged with a series of proteins to make an **RNA-induced silencing complex** (**RISC**). Such small regulatory RNAs can bind to the 3′ UTR of messages and inhibit their translation. In some cases (especially when the binding of the miRNA to the 3′ UTR is tight), the site is cleaved. More usually, however, several RISCs attach to sites on the 3′ UTR and prevent the message from being translated (see Bartel 2004; He and Hannon 2004).

The abundance of microRNAs and their apparent conservation among flies, nematodes, vertebrates, and even plants suggest that such RNA regulation is a previously unrecognized but potentially very important means of regulating gene expression. This hidden layer of gene regulation parallels the better known protein-level gene control mechanisms, and it may be just as important in regulating cell fate. Recent studies have shown that microRNAs are involved in mammalian heart and blood cell differentiation. During mouse heart development, the microRNA *miR1* can repress the messages encoding the Hand2 transcription factor (Zhao et al. 2005). This transcription factor is critical in the proliferation of ventricle heart muscle cells, and *miR1* may control the balance between ventricle growth and differentiation. The *miR181* miRNA is essential for committing progenitor cells to differentiate into B lymphocytes, and ectopic expression of *miR181* in mice causes a preponderance of B lymphocytes (Figure 2.36; Chen et al. 2004).

MicroRNAs are also used to "clean up" and fine-tune the level of gene products. We mentioned those maternal RNAs that allow early development to occur. How does the embryo get rid of maternal RNAs once they have been used and the embryonic cells are making their own mRNAs? In zebrafish, this cleanup operation is assigned to microRNAs such as *miR430*. This is one of the first genes transcribed by the fish embryonic cells, and there are about 90 copies of this gene in the zebrafish genome. So the level

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

SIDELIGHTS & SPECULATIONS

microRNAs in Transcriptional Gene Regulation

In addition to its role in the translational regulation of gene expression, microRNAs also appear to be able to silence the transcription of certain genes. Such genes are often located in the heterochromatin—that region of the genome where the DNA is tightly coiled and transcription is inhibited by the packed nucleosomes. Volpe and colleagues (2003) discovered that if they deleted the genes in yeast encoding the appropriate RNases or RISC proteins, the heterochromatin around the centromeres became unpacked, the histones in this region lost their inhibitory methylation, and the centromeric heterochromatin started making RNA. Similar phenomena were seen when these proteins were mutated in *Drosophila* (Pal-Bhadra et al. 2004). Indeed, in *Drosophila*, the *Suppressor-of-stellate* gene on the Y chromosome makes a microRNA that represses the transcription of the *stellate* gene on the X chromosome (Gvozdev et al. 2003). This is important for dosage regulation of the X chromosomes in *Drosophila*.

It appears that microRNAs are able to bind to the nuclear RNA as it is being transcribed, and form a complex with the methylating and deacetylating enzymes, thus repressing the gene (Kato et al. 2005; Schramke et al. 2005). If synthetic microRNA made complementary to specific promoters is added to cultured human cells, that microRNA is able to induce that promoter's DNA to become methylated. Lysine 9 on histone H3 also becomes methylated around the promoter, and transcription from that gene stops (Kawasaki and Taira 2004; Morris et al. 2004).

This appears to be the mechanism by which NRSF (see page 47) functions. NRSF prevents gene expression in nonneural cells by repressing microRNAs that would otherwise recruit histone acetyltransferases to activate genes that promote neuron production. In the presence of NRSF, these miRNAs are not present, and so histone deacetylases and methyltransferases are recruited to the chromatin instead. The resulting methylation produces conglomerations of nucleosomes linked together by heterochromatin protein-1 (HP1), thereby stabilizing the conglomerate and preventing transcription of the neuron-promoting genes "hidden" within it (Ooi and Wood 2007; Yoo et al. 2009.) A single silencer protein bound to the DNA can prevent the gene's expression.

Thus, microRNA directed against the 3′ end of mRNA may be able to shut down gene expression on the translational level, while microRNA directed at the promoters of genes may be able to block gene expression at the transcriptional level. The therapeutic value of these RNAs in cancer therapy is just beginning to be explored (see Gaur and Rossi 2006).

X chromosome inactivation in mammals is also directed by small noncoding RNAs, albeit not the canonical microRNAs. Although DNA methylation is responsible for keeping one of the two X chromosomes inactive, the choice of which X chromosome to activate arises from the physical interactions between the two X chromosomes and their expressing several small, noncoding RNAs. The mechanisms of differential expression of small RNAs between the two X chromosomes is under intensive investigation (Augui et al. 2007; Migeon 2007; Ogawa et al. 2008).

of *miR430* goes up very rapidly. This microRNA has hundreds of targets (about 40% of the maternal RNA types), and when it binds to the 3′ UTR of these target mRNAs, these mRNAs lose their polyA tails and are degraded (Giraldez et al. 2006). Slightly later in development, this same microRNA is used in the fish embryo to fine-tune the expression of *Nodal* mRNA (Choi et al. 2007). The consequence of this latter use of *miR430* is the determination of how many cells become committed to the endoderm and how many become committed to be mesoderm.

Although the microRNA is usually about 22 bases long, it recognizes its target primarily through a "seed" region

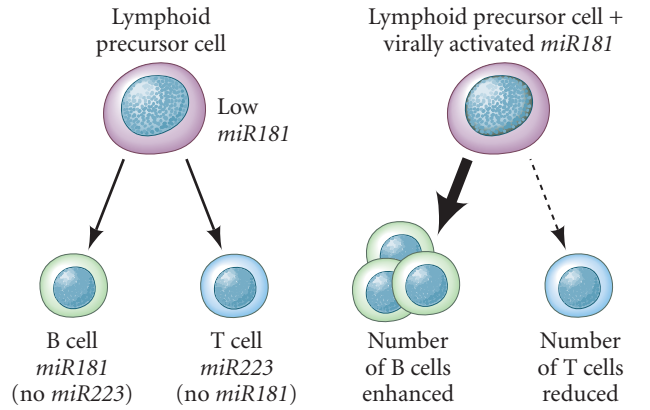


FIGURE 2.36 The lymphoid precursor cell can generate B cells (lymphocytes that make antibodies) or T cells (lymphocytes that kill virally infected cells). This differentiation depends on the organ in which they reside. The regulation of the lineage pathway is controlled in part by levels of the microRNA *miR181*. The lymphocyte precursor cell has little *miR181*. A B cell has high levels of *miR181*, whereas T cells do not appear to have any. If lymphocyte precursor cells are virally transfected with *miR181*, they preferentially generate B cells at the expense of T cells.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

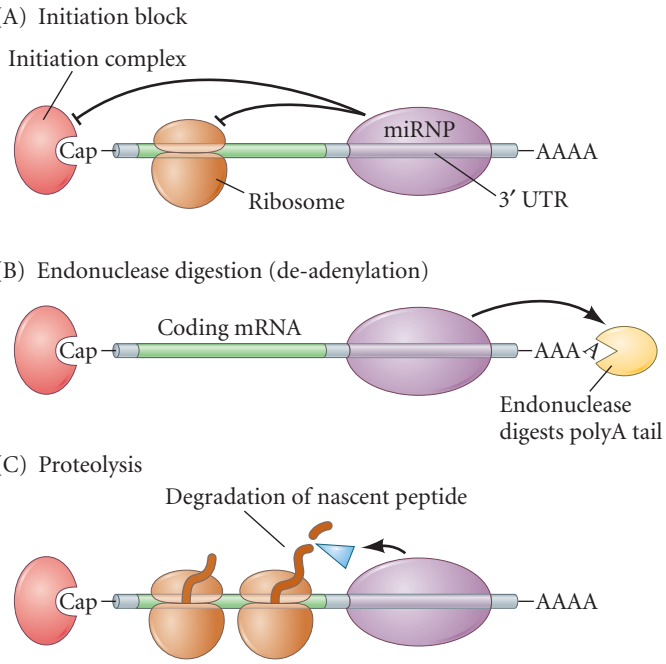


FIGURE 2.37 The miRNA complex, including numerous proteins that bind to the miRNA (miRNP), can block translation in several ways. These include (A) blocking the binding of the mRNA to initiation factors or ribosomes; (B) recruiting endonucleases to chew away the polyA tail of the mRNA, thereby causing its destruction; and (C) recruiting protein-digesting enzymes that destroy the nascent protein. (After Filipowicz et al. 2008.)

of about 5 bases in the 5′ end of the microRNA (usually at positions 2–7). This seed region recognizes targets in the 3′ UTR of the message. What happens, then, if an mRNA has a mutated 3′ UTR? Such a mutation appears to have given rise to the Texel sheep, a breed with a large and well-defined musculature that is the dominant meat-producing sheep in Europe. We have already seen that a mutation in the *myostatin* gene that prevents the proper splicing of the nRNA can produce a large-muscled phenotype. Another way of reducing the levels of myostatin involves a mutation in its 3′UTR sequence (see Figure 2.30). Genetic techniques mapped the basis of the sheep's meaty phenotype to the *myostatin* gene. In the Texel breed, there has been a G-to-A transition in the 3′ UTR of the gene for myostatin, creating a target for the *mir1* and *mir206* microRNAs that are abundant in skeletal muscle (Clop et al. 2006). This mutation causes the depletion of myostatin messages and the increase in muscle mass characteristic of these sheep.

The binding of microRNAs to the 3′ UTR can regulate translation in several ways (Figure 2.37; Filipowicz et al. 2008). First, they can block initiation of translation, preventing the binding of initiation factors or ribosomes. Second, they can recruit endonucleases that digest the mRNA, usually starting with the polyA tail. In a third mechanism, they allow translation to be initiated, but recruit proteolytic enzymes that digest the protein as it is being made. It is also possible that some microRNAs use more than one method, and it has been proposed (Mathonnet et al. 2007) that the microRNAs may first inhibit translation initiation and then consolidate mRNA silencing by causing the digestion of the message.

Control of RNA expression by cytoplasmic localization

Not only is the time of mRNA translation regulated, but so is the place of RNA expression. A majority of mRNAs (about 70% in *Drosophila* embryos) are localized to specific places in the cell (Lécuyer et al. 2007). Just like the selective repression of mRNA translation, the selective localization of messages is often accomplished through their 3′ UTRs. There are three major mechanisms for the localization of an mRNA (Figure 2.38; see Palacios 2007):

- *Diffusion and local anchoring*. Messenger RNAs such as *nanos* diffuse freely in the cytoplasm. However, when they diffuse to the posterior pole of the *Drosophila* oocyte, they are trapped there by proteins that reside particularly in these regions. These proteins also activate the mRNA, allowing it to be translated.
- *Localized protection.* Messenger RNAs such as those encoding the *Drosophila* heat shock protein hsp83 (which helps protect the embryos from thermal extremes) also float freely in the cytoplasm. Like *nanos* mRNA, *hsp83* accumulates at the posterior pole, but its mechanism for getting there is different. Throughout the embryo, the protein is degraded. However, proteins at the posterior pole protect the *hsp83* mRNA from being destroyed.
- *Active transport along the cytoskeleton.* This is probably the most widely used mechanism for mRNA localization. Here, the 3′ UTR of the mRNA is recognized by proteins that can bind these messages to "motor proteins" that travel along the cytoskeleton to their final destination. These motor proteins are usually ATPases such as dynein or kinesin that split ATP for their motive force. For instance, in *Drosophila* oocytes, the *bicoid* messages (which instruct the formation of the head) are localized to one end of the oocyte. The 3′ UTR of *bicoid* mRNA allows its message to bind to the microtubules through its association with two other proteins (Swallow and Staufen). If the *bicoid* 3′ UTR is attached to some other message, that mRNA will also be bound to the anterior pole of the oocyte (Driever and Nüsslein-Volhard 1988a,b; Ferrandon et al. 1994).

The 3′ UTR of the *bicoid* message binds the Staufen protein that connects it to dynein. Dynein travels along the microtubules in the "minus" direction, that is, toward the site where microtubules begin. In this way, the *bicoid* mRNA is localized to the future anterior part of the oocyte. Other mRNAs, such as the *Oskar* message, in contrast, appear to

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

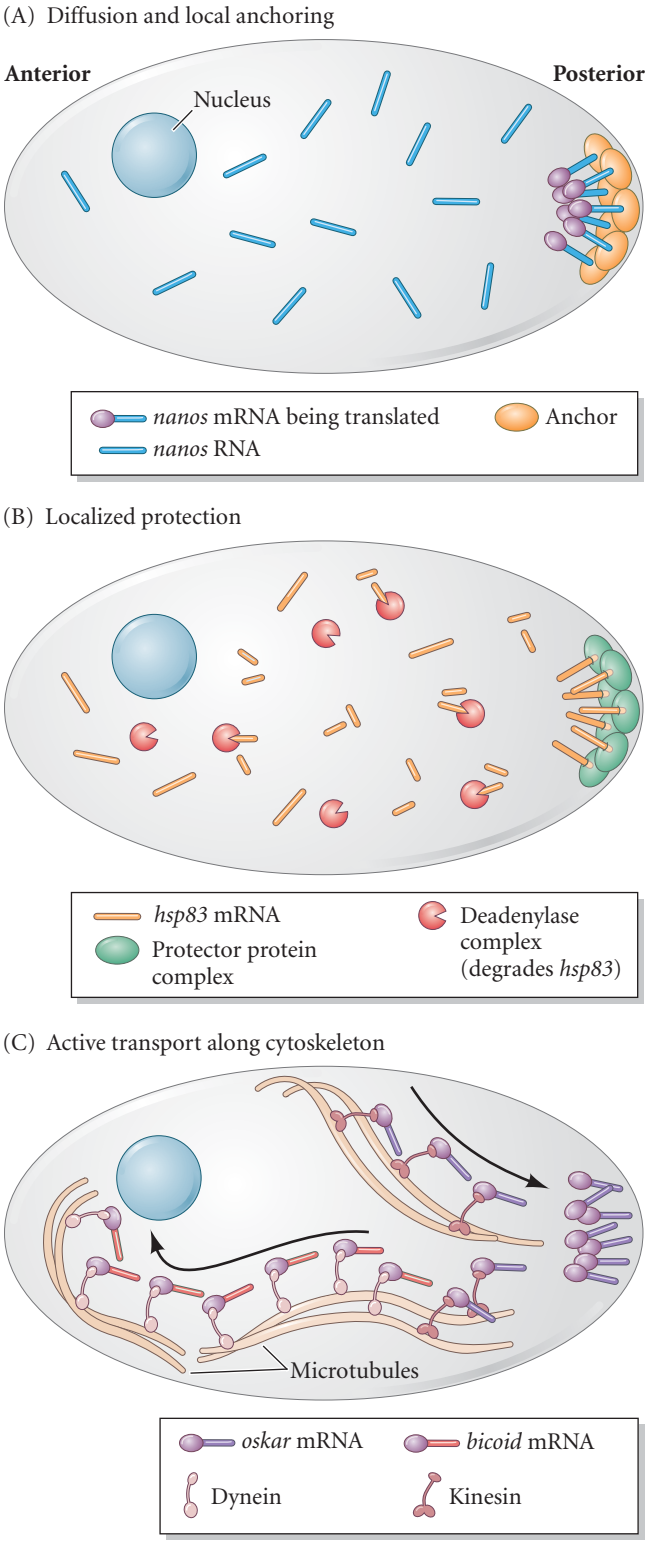


FIGURE 2.38 Localization of mRNAs. (A) Diffusion and local anchoring. *Nanos* mRNA diffuses through the *Drosophila* egg and is bound (in part by the Oskar protein, whose message is described in the text) at the posterior end of the oocyte. This anchoring allows the *nanos* mRNA to be translated. (B) Localized protection. The mRNA for *Drosophila* heat shock protein (hsp83) will be degraded unless it binds to a protector protein (in this case, also at the posterior terminal of the oocyte). (C) Active transport on the cytoskeleton, causing the accumulation of mRNA at a particular site. Here, *bicoid* mRNA is transported to the anterior of the oocyte by dynein and kinesin motor proteins. Meanwhile, *Oskar* mRNA is brought to the posterior pole by transport along microtubules by kinesin ATPases. (After Palacios 2007.)

bind to the kinesin motor protein, and it is taken toward the "plus" end of the microtubules, at the tip of their assembly. It is thereby taken to the posterior end of the *Drosophila* oocyte. Once transported to their destinations, mRNAs often bind to other cytoskeletal proteins (such as actin microfilaments).

Stored mRNAs in brain cells

One of the most important areas of local translational regulation may be in the brain. The storage of long-term memory requires new protein synthesis, and the local translation of mRNAs in the dendrites of brain neurons has been proposed as a control point for increasing the strength of synaptic connections (Martin 2000; Klann et al. 2004; Wang and Tiedge 2004). The ability to increase the strength of the connections between neurons is critical in forming the original architecture of the brain and also in the ability to learn. Indeed, in recent studies of mice, Kelleher and colleagues (2004) have shown that neuronal activity-dependent memory storage depends on the activation of eIF4E and other components of protein synthesis.

Several mRNAs appear to be transported along the cytoskeleton to the dendrites of neurons (the "receiving portion" of the neuron, where synapse connections are formed with the other neurons). These messages include those mRNAs encoding receptors for neurotransmitters (needed to transmit the signals from one neuron to another); activity-regulated enzymes; and the cytoskeletal components needed to build a synapse (Figure 2.39). As we will see in later chapters, one of the proteins responsible for constructing specific synapses is **brain-derived neurotrophic factor**, or **BDNF**. BDNF regulates neural activity and appears to be critical for new synapse formation. Takei and colleagues (2004) have shown that BDNF induces local translation of these neural messages in the dendrites.

Another indication of the importance of dendritic mRNA translation comes from studies of a leading cause of human mental retardation, fragile X syndrome. Fragile X syndrome is caused by loss-of-function mutations in the X-linked *FMR1* gene. The FMR1 protein appears to prevent the translation of several mRNAs that are being transported to the dendrites along microtubules in response to stimulation by glutamic acid (Dictenberg et al. 2008; Wang et al. 2008b). In the absence of functional FMR1, these mRNAs are expressed in the wrong amounts, leading to signaling abnormalities that are believed to cause the prob-

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

FIGURE 2.39 A brain-specific RNA in a cultured mammalian neuron. BC1 RNA (stained white) appears to be clustered at specific sites in the neuron (stained light blue), especially in the dendrites. (From Wang and Tiedge 2004, photograph courtesy of the authors.)

lems in cognition and learning. Thus, translational regulation in neurons might be important not only for their initial development but also for their continued ability to learn and change.

Posttranslational regulation of gene expression

When a protein is synthesized, the story is still not over. Once a protein is made, it becomes part of a larger level of organization. For instance, it may become part of the structural framework of the cell, or it may become involved in one of the myriad enzymatic pathways for the synthesis or breakdown of cellular metabolites. In any case, the individual protein is now part of a complex "ecosystem" that integrates it into a relationship with numerous other proteins. Thus, several changes can still take place that determine whether or not the protein will be active.

Some newly synthesized proteins are inactive without the cleaving away of certain inhibitory sections. This is what happens when insulin is made from its larger protein precursor. Some proteins must be "addressed" to their specific intracellular destinations in order to function. Proteins are often sequestered in certain regions of the cell, such as membranes, lysosomes, nuclei, or mitochondria. Some proteins need to assemble with other proteins in order to form a functional unit. The hemoglobin protein, the microtubule, and the ribosome are all examples of numerous proteins joining together to form a functional unit. And some proteins are not active unless they bind an ion (such as Ca^{2+}), or are modified by the covalent addition of a phosphate or acetate group. The importance of this last type of protein modification will become obvious in Chapter 3, since many of the critical proteins in embryonic cells just sit there until some signal activates them.

Snapshot Summary: *Developmental Genetics*

1. Differential gene expression from genetically identical nuclei creates different cell types. Differential gene expression can occur at the levels of gene transcription, nuclear RNA processing, mRNA translation, and protein modification. Notice that RNA processing and export occur while the RNA is still being transcribed from the gene.
2. Genes are usually repressed, and activating a gene often means inhibiting its repressor. This fact leads to thinking in double and triple negatives: Activation is often the inhibition of the inhibitor; repression is the inhibition of the inhibitor of the inhibitor.
3. Eukaryotic genes contain promoter sequences to which RNA polymerase can bind to initiate transcription. To accomplish this, the eukaryotic RNA polymerases are bound by a series of proteins called transcription-associated factors, or TAFs.
4. Eukaryotic genes expressed in specific cell types contain enhancer sequences that regulate their transcription in time and space.
5. Specific transcription factors can recognize specific sequences of DNA in the promoter and enhancer

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

regions. These proteins activate or repress transcription from the genes to which they have bound.

6. Enhancers work in a combinatorial fashion. The binding of several transcription factors can act to promote or inhibit transcription from a certain promoter. In some cases transcription is activated only if both factor A and factor B are present; in other cases, transcription is activated if either factor A or factor B is present.
7. A gene encoding a transcription factor can keep itself activated if the transcription factor it encodes also activates its own promoter. Thus, a transcription factor gene can have one set of enhancer sequences to initiate its activation and a second set of enhancer sequences (which bind the encoded transcription factor) to maintain its activation.
8. Often, the same transcription factors that are used during the differentiation of a particular cell type are also used to activate the genes for that cell type's specific products.
9. Enhancers can act as silencers to suppress the transcription of a gene in inappropriate cell types.
10. Transcription factors act in different ways to regulate RNA synthesis. Some transcription factors stabilize RNA polymerase binding to the DNA; some disrupt nucleosomes, increasing the efficiency of transcription.
11. Transcription correlates with a lack of methylation on the promoter and enhancer regions of genes. Methylation differences can account for examples of genomic imprinting, wherein a gene transmitted through the sperm is expressed differently than the same gene transmitted through the egg.
12. Dosage compensation enables the X chromosome-derived products of males (which have one X chromosome per cell in fruit flies and mammals) to equal the X chromosome-derived products of females (which have two X chromosomes per cell). This compensation is accomplished at the level of transcription, either by accelerating transcription from the lone X chromosome in males (*Drosophila*), decreasing the level of transcription from each X chromosome by 50% (*C. elegans*), or by inactivating a large portion of one of the two X chromosomes in females (mammals).
13. Differential nuclear RNA selection can allow certain transcripts to enter the cytoplasm and be translated while preventing other transcripts from leaving the nucleus.
14. Differential RNA splicing can create a family of related proteins by causing different regions of the nRNA to be read as exons or introns. What is an exon in one set of circumstances may be an intron in another.
15. Some messages are translated only at certain times. The oocyte, in particular, uses translational regulation to set aside certain messages that are transcribed during egg development but used only after the egg is fertilized. This activation is often accomplished either by the removal of inhibitory proteins or by the polyadenylation of the message.
16. MicroRNAs can act as translational inhibitors, binding to the 3′ UTR of the RNA.
17. Many mRNAs are localized to particular regions of the oocyte or other cells. This localization appears to be regulated by the 3′ UTR of the mRNA.

For Further Reading

Complete bibliographical citations for all literature cited in this chapter can be found at the free access website **www.devbio.com**

Clop, A. and 16 others. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genet.* 38: 813–818.

Davidson, E. H. 2006. *The Regulatory Genome*. Academic Press, New York.

Migeon, B. R. 2007. *Females Are Mosaics: X Inactivation and Sex Differences in Disease*. Oxford University Press, New York.

Palacios, I. M. 2007. How does an mRNA find its way? Intracellular localization of transcripts. *Sem. Cell Dev. Biol.* 163–170.

Strahl, B. D. and C. D. Allis. 2000. The language of covalent histone modifications. *Nature* 403: 41–45.

Wilmut, I., K. Campbell and C. Tudge. 2001. *The Second Creation: Dolly and the Age of Biological Control.* Harvard University Press, Cambridge, MA.

Zhou, Q., J. Brown, A. Kanarek, J. Rajagopal and D. A. Melton. 2008. In vivo reprogramming of adult pancreatic exocrine cells to β cells. *Nature* 455: 627–632.

Zinzen, R. P., C. Girardot, J. Gagneur, M. Braun and E. E. Furlong. 2009. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* 462: 65–70.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

Go Online

WEBSITE 2.1 **Does the genome or the cytoplasm direct development?** The geneticists versus the embryologists. Geneticists were certain that genes controlled development, whereas embryologists generally favored the cytoplasm. Both sides had excellent evidence for their positions.

WEBSITE 2.2 **The origins of developmental genetics.** The first hypotheses for differential gene expression came from C. H. Waddington, Salome Gluecksohn-Waelsch, and other scientists who understood both embryology and genetics.

WEBSITE 2.3 **Techniques of DNA analysis.** The entries of this website describe crucial laboratory skill including gene cloning, DNA sequencing, Southern blotting, "knock-outs" of specific genes, enhancer traps, and identification of methylated sites.

WEBSITE 2.4 **Techniques of RNA analysis.** Techniques described here include northern blots, RT-PCR, in situ hybridization, microarray technology, antisense RNA, interference RNA, Cre-lox analysis, and ChIP-on-Chip.

WEBSITE 2.5 **Bioinformatics.** This entry provides links to various free websites with tools that enable researchers to compare DNA sequences and specific genomes with the aim of further illuminating the various mechanisms of gene regulation.

WEBSITE 2.6 **Cloning and nuclear equivalence.** The several entries here address the issues of cloning and whether or not the entire genome is the same in each cell of the body. As it turns out, lymphocytes make new genes during their development and their genomes are not identical.

WEBSITE 2.7 **Silencing large blocks of chromatin.** The inactivation or the elimination of entire chromosomes is not uncommon among invertebrates and is sometimes used as a mechanism of sex determination. Moreover, among mammals, random X chromosome inactivation may provide females with health benefits—as long as the process occurs flawlessly.

WEBSITE 2.8 **So you think you know what a gene is?** Different scientists have different definitions, and nature has given us some problematic examples of DNA sequences that may or may not be considered genes.

UNCORRECTED PAGE PROOFS © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.