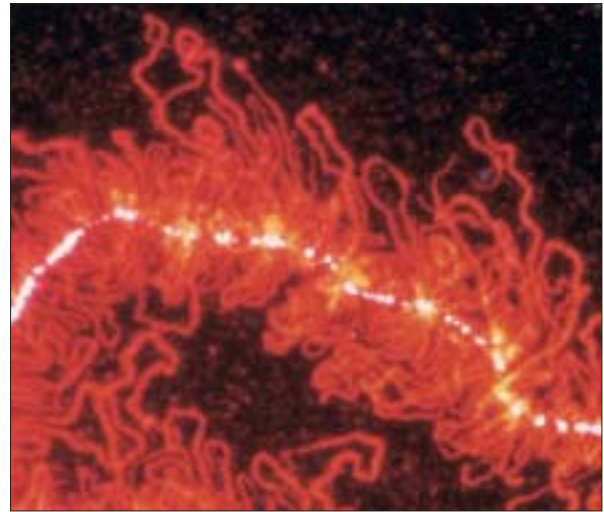


12

POST-TRANSCRIPTIONAL GENE CONTROL AND NUCLEAR TRANSPORT



Portion of a “lampbrush chromosome” from an oocyte of the newt *Nophthalmus viridescens*; hnRNP protein associated with nascent RNA transcripts fluoresces red after staining with a monoclonal antibody. [Courtesy of M. Roth and J. Gall.]

In the previous chapter, we saw that most genes are regulated at the first step in gene expression, namely, the initiation of transcription. However, once transcription has been initiated, synthesis of the encoded RNA requires that RNA polymerase transcribe the entire gene and not terminate prematurely. Moreover, the initial **primary transcripts** produced from eukaryotic genes must undergo various processing reactions to yield the corresponding functional RNAs. Once formed in the nucleus, mature, functional RNAs are transported to the cytoplasm as components of ribonucleoproteins. Both processing of RNAs and their export from the nucleus offer opportunities for further regulating gene expression at stages subsequent to transcription initiation. Additional control of gene expression can occur in the cytoplasm. In the case of protein-coding genes, for instance, the amount of protein produced depends on the stability of the corresponding mRNAs in the cytoplasm and the rate of their translation. In addition, the cellular locations of some mRNAs are regulated, so that newly synthesized protein is concentrated where it is needed.

We refer to all the mechanisms that regulate gene expression following transcription as *post-transcriptional gene control* (Figure 12-1). Although these mechanisms generally play a smaller overall role in **gene control** than activation or repression of transcription initiation, they are critical in regulating expression of some genes. In this chapter, we consider the events that occur in the processing of mRNA following transcription initiation and the various mechanisms that are known to regulate these events. In the last section, we briefly discuss the processing of primary transcripts produced from genes encoding rRNAs and tRNAs.

12.1 Processing of Eukaryotic Pre-mRNA

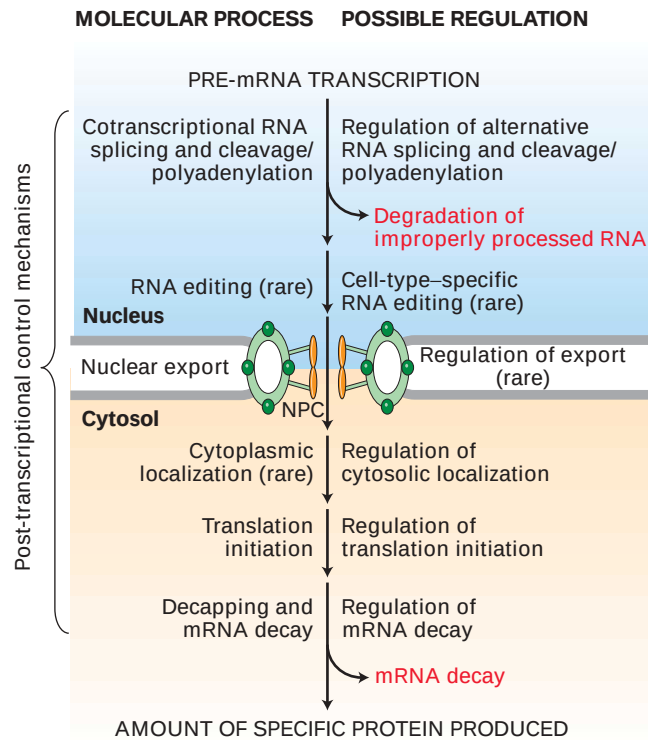
In this section, we take a closer look at how eukaryotic cells convert the initial primary transcript synthesized by RNA polymerase II into a functional mRNA. Three major events occur during the process: *5' capping*, *3' cleavage/polyadenylation*, and *RNA splicing* (Figure 12-2). Processing occurs in the nucleus as the nascent mRNA precursor is being transcribed, and the functional mRNA produced is transported to the cytoplasm by mechanisms discussed later.

The 5' Cap Is Added to Nascent RNAs Shortly After Initiation by RNA Polymerase II

After nascent RNA molecules produced by RNA polymerase II reach a length of 25–30 nucleotides, 7-methylguanosine and the other components of the *5' cap* found on eukaryotic

OUTLINE

- 12.1 Processing of Eukaryotic Pre-mRNA
- 12.2 Regulation of Pre-mRNA Processing
- 12.3 Macromolecular Transport Across the Nuclear Envelope
- 12.4 Cytoplasmic Mechanisms of Post-transcriptional Control
- 12.5 Processing of rRNA and tRNA



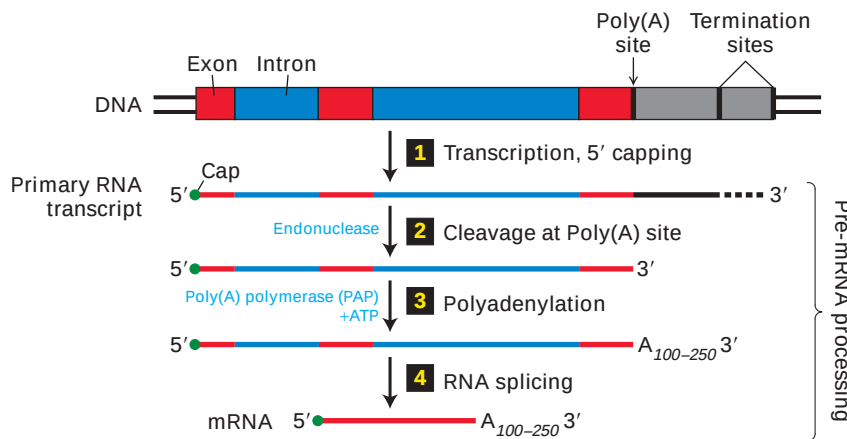
▲ **FIGURE 12-1 Overview of post-transcriptional control of protein-coding genes.** Control may be exerted as a primary transcript is processed in the nucleus, during export of an mRNA to the cytoplasm, or in the cytoplasm. Any one gene would likely be regulated by only one or a few of the possible control mechanisms.

mRNAs are added to their 5' end (see Figure 4-13). This initial step in RNA processing is catalyzed by a dimeric capping enzyme, which associates with the phosphorylated *carboxyl-terminal domain* (CTD) of RNA polymerase II. Recall that the CTD becomes phosphorylated during transcription initiation (see Figure 11-27). Because the capping enzyme does not associate with polymerase I or III, which do not contain a CTD, capping is specific for transcripts produced by RNA polymerase II.

One subunit of the capping enzyme removes the γ phosphate from the 5' end of the nascent RNA emerging from the surface of an RNA polymerase II. Another domain of this subunit transfers the GMP moiety from GTP to the 5'-diphosphate of the nascent transcript, creating the unusual guanosine 5'-5'-triphosphate structure. In the final steps, separate enzymes transfer methyl groups from S-adenosyl-methionine to the N₇ position of the guanine and the 2' oxygens of riboses at the 5' end of the nascent RNA.

Pre-mRNAs Are Associated with hnRNP Proteins Containing Conserved RNA-Binding Domains

Nascent RNA transcripts from protein-coding genes and mRNA processing intermediates, collectively referred to as **pre-mRNA**, do not exist as free RNA molecules in the nuclei of eukaryotic cells. From the time nascent transcripts first emerge from RNA polymerase II until mature mRNAs are transported into the cytoplasm, the RNA molecules are associated with an abundant set of nuclear proteins. These



▲ **FIGURE 12-2 Overview of mRNA processing in eukaryotes.** Shortly after RNA polymerase II initiates transcription at the first nucleotide of the first exon of a gene, the 5' end of the nascent RNA is capped with 7-methylguanylate (step **1**). Transcription by RNA polymerase II terminates at any one of multiple termination sites downstream from the poly(A) site, which is located at the 3' end of the final exon. After the primary transcript is cleaved at the poly(A) site (step **2**),

a string of adenosine (A) residues is added (step **3**). The poly(A) tail contains ≈ 250 A residues in mammals, ≈ 150 in insects, and ≈ 100 in yeasts. For short primary transcripts with few introns, splicing (step **4**) usually follows cleavage and polyadenylation, as shown. For large genes with multiple introns, introns often are spliced out of the nascent RNA during its transcription, i.e., before transcription of the gene is complete. Note that the 5' cap and sequence adjacent to the poly(A) tail are retained in mature mRNAs.

proteins are the major protein components of *heterogeneous ribonucleoprotein particles (hnRNPs)*, which contain *heterogeneous nuclear RNA (hnRNA)*, a collective term referring to pre-mRNA and other nuclear RNAs of various sizes. The proteins in these ribonucleoprotein particles can be dramatically visualized with fluorescent-labeled monoclonal antibodies (see the chapter opening figure).

Researchers identified hnRNP proteins by first exposing cultured cells to high-dose UV irradiation, which causes covalent cross-links to form between RNA bases and closely associated proteins. Chromatography of nuclear extracts from treated cells on an oligo-dT cellulose column, which binds RNAs with a poly(A) tail, was used to recover the proteins that had become cross-linked to nuclear polyadenylated RNA. Subsequent treatment of cell extracts from unirradiated cells with monoclonal antibodies specific for the major proteins identified by this cross-linking technique revealed a complex set of abundant hnRNP proteins ranging in size from 34 to 120 kDa.

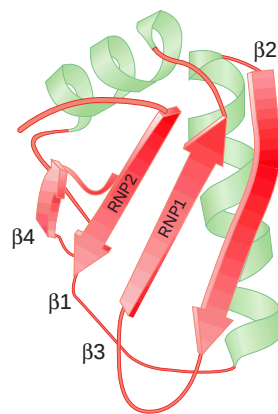
Like transcription factors, most hnRNP proteins have a modular structure. They contain one or more RNA-binding domains and at least one other domain that is thought to interact with other proteins. Several different RNA-binding motifs have been identified by constructing deletions of hnRNP proteins and testing their ability to bind RNA.

Conserved RNA-Binding Motifs The *RNA recognition motif (RRM)*, also called the RNP motif and the RNA-

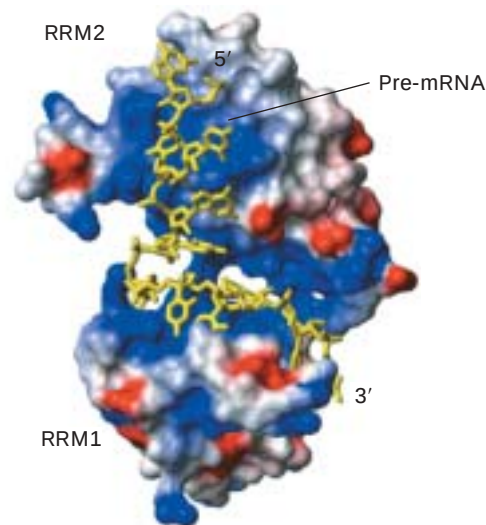
binding domain (RBD), is the most common RNA-binding domain in hnRNP proteins. This ≈80-residue domain, which occurs in many other RNA-binding proteins, contains two highly conserved sequences (RNP1 and RNP2) that allow the motif to be recognized in newly sequenced genes. X-ray crystallographic analysis has shown that the RRM domain consists of a four-stranded β sheet flanked on one side by two α helices. The conserved RNP1 and RNP2 sequences lie side by side on the two central β strands, and their side chains make multiple contacts with a single-stranded region of RNA (Figure 12-3). The single-stranded RNA lies across the surface of the β sheet, with the central RNP1 and RNP2 strands forming a positively charged surface that interacts with the negatively charged RNA phosphates.

The *RGG box*, another RNA-binding motif found in hnRNP proteins, contains five Arg-Gly-Gly (RGG) repeats with several interspersed aromatic amino acids. Although the structure of this motif has not yet been determined, its arginine-rich nature is similar to the RNA-binding domains of the HIV Tat protein. The 45-residue *KH motif* is found in the hnRNP K protein and several other RNA-binding proteins; commonly two or more copies of the KH motif are interspersed with RGG repeats. The three-dimensional structure of representative KH domains is similar to that of the RRM domain but smaller, consisting of a three-stranded β sheet supported from one side by two α helices. RNA binds to the KH motif by interacting with a hydrophobic surface formed by the α helices and one

(a) RNA recognition motif (RRM)



(b) Sex-lethal RRM domains



▲ FIGURE 12-3 Structure of the RRM domain and its interaction with RNA. (a) Diagram of the RRM domain showing the two α helices and four β strands that characterize this motif. The conserved RNP1 and RNP2 regions are located in the two central β strands. (b) Surface representation of the two RRM domains in *Drosophila* Sex-lethal (Sxl) protein, which binds a nine-base sequence in *transformer* pre-mRNA (green). The two RRMs are oriented like the two parts of an open pair of castanets, with

the β sheet of RRM1 facing upward and the β sheet of RRM2 facing downward. Positively charged regions in Sxl protein are shown in shades of blue; negatively charged regions, in shades of red. The pre-mRNA is bound to the surfaces of the positively charged β sheets, making most of its contacts with the RNP1 and RNP2 regions of each RRM. [Part (a) adapted from K. Nagai et al., 1995, *Trends Biochem. Sci.* **20**: 235; part (b) after N. Harada et al., 1999, *Nature* **398**:579.]

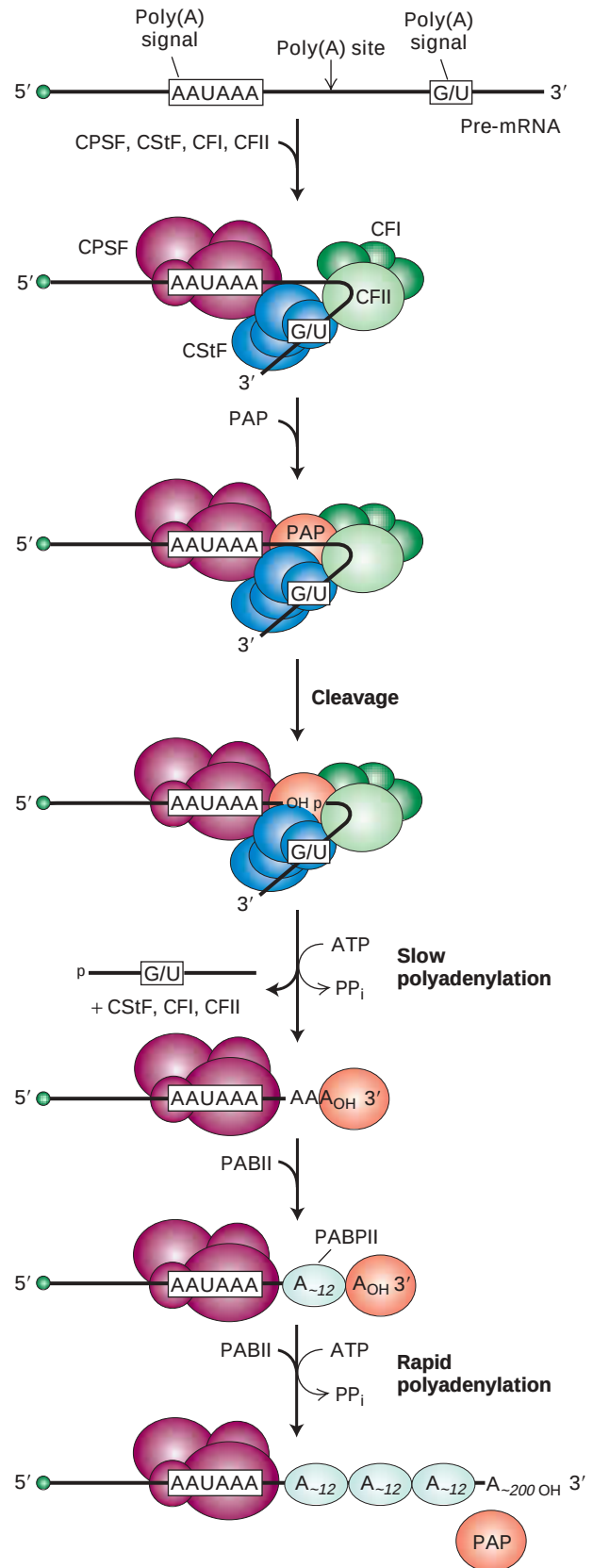
β -strand. Thus, despite the similarity in their structures, RRM and KH domains interact differently with RNA.

Functions of hnRNP Proteins The association of pre-mRNAs with hnRNP proteins prevents formation of short secondary structures dependent on base-pairing of complementary regions, thereby making the pre-mRNAs accessible for interaction with other RNA molecules or proteins. Pre-mRNAs associated with hnRNP proteins present a more uniform substrate for further processing steps than would free, unbound pre-mRNAs, each type of which forms a unique secondary structure dependent on its specific sequence.

Binding studies with purified hnRNP proteins suggest that different hnRNP proteins associate with different regions of a newly made pre-mRNA molecule. For example, the hnRNP proteins A1, C, and D bind preferentially to the pyrimidine-rich sequences at the 3' ends of introns (see later discussion). This observation suggests that some hnRNP proteins may interact with the RNA sequences that specify RNA splicing or cleavage/polyadenylation and contribute to the structure recognized by RNA-processing factors. Finally, cell-fusion experiments have shown that some hnRNP proteins remain localized in the nucleus, whereas others cycle in and out of the cytoplasm, suggesting that they function in the transport of mRNA. We discuss the role of these proteins in nuclear transport in Section 12.3.

3' Cleavage and Polyadenylation of Pre-mRNAs Are Tightly Coupled

In eukaryotic cells, all mRNAs, except histone mRNAs, have a 3' *poly(A) tail*. Early studies of pulse-labeled adenovirus and SV40 RNA demonstrated that the viral primary transcripts extend beyond the site in the viral mRNAs from which the poly(A) tail extends. These results suggested that A residues are added to a 3' hydroxyl generated by endonucleolytic cleavage of a longer transcript, but the predicted downstream RNA fragments never were detected *in vivo*, presumably because of their rapid degradation. That cleavage of a primary transcript precedes its polyadenylation was firmly established by detection of both predicted cleavage



► **FIGURE 12-4 Model for cleavage and polyadenylation of pre-mRNAs in mammalian cells.** Cleavage and polyadenylation specificity factor (CPSF) binds to the upstream AAUAAA poly(A) signal. CSTF interacts with a downstream GU- or U-rich sequence and with bound CPSF, forming a loop in the RNA; binding of CFI and CFII help stabilize the complex. Binding of poly(A) polymerase (PAP) then stimulates cleavage at a poly(A) site, which usually is 10–35 nucleotides 3' of the upstream poly(A) signal. The cleavage factors are released, as is the downstream RNA cleavage product, which is rapidly degraded. Bound PAP then adds ≈ 12 A residues at a slow rate to the 3'-hydroxyl group generated by the cleavage reaction. Binding of poly(A)-binding protein II (PABPII) to the initial short poly(A) tail accelerates the rate of addition by PAP. After 200–250 A residues have been added, PABPII signals PAP to stop polymerization.

products in *in vitro* processing reactions performed with nuclear extracts of HeLa cells.

Early sequencing of cDNA clones from animal cells showed that nearly all mRNAs contain the sequence AAUAAA 10–35 nucleotides *upstream* from the poly(A) tail. Polyadenylation of RNA transcripts is virtually eliminated when the corresponding sequence in the template DNA is mutated to any other sequence except one encoding a closely related sequence (AUUAAA). The unprocessed RNA transcripts produced from such mutant templates do not accumulate in nuclei, but are rapidly degraded. Further mutagenesis studies revealed that a second signal *downstream* from the cleavage site is required for efficient cleavage and polyadenylation of most pre-mRNAs in animal cells. This downstream signal is not a specific sequence but rather a GU-rich or simply a U-rich region within ≈ 50 nucleotides of the cleavage site.

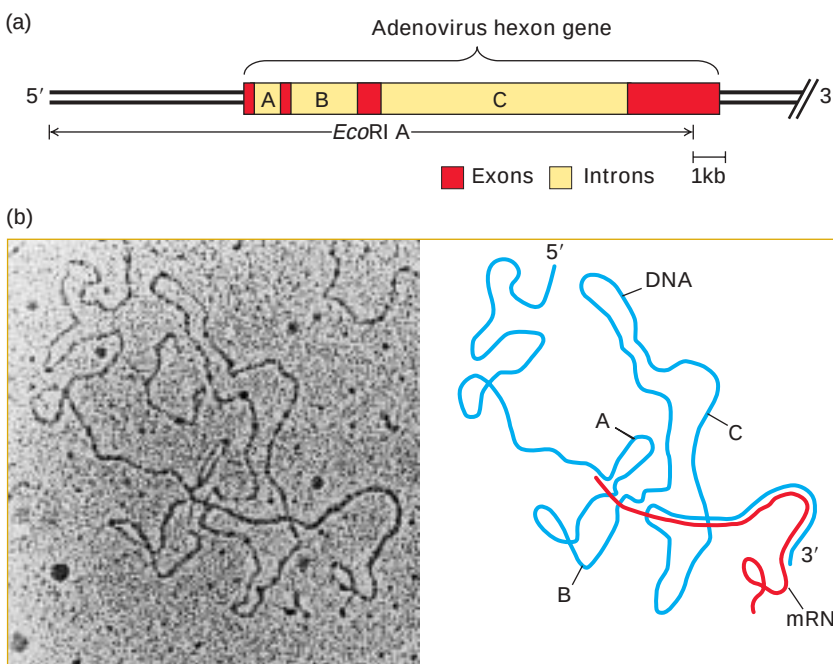
Identification and purification of the proteins required for cleavage and polyadenylation of pre-mRNA have led to the model shown in Figure 12-4. According to this model, a 360-kDa *cleavage and polyadenylation specificity factor* (CPSF), composed of four different polypeptides, first forms an unstable complex with the upstream AAUAAA poly(A) signal. Then at least three additional proteins bind to the CPSF-RNA complex: a 200-kDa heterotrimer called *cleavage stimulatory factor* (CStF), which interacts with the G/U-rich sequence; a 150-kDa heterotrimer called *cleavage factor I* (CFI); and a second, poorly characterized cleavage factor (CFII). Finally, a *poly(A) polymerase* (PAP) binds to the complex *before* cleavage can occur. This requirement for PAP binding links cleavage and polyadenylation, so that the free 3' end generated is rapidly polyadenylated.

Assembly of the large, multiprotein **cleavage/polyadenylation complex** around the AU-rich poly(A) signal in a pre-mRNA is analogous in many ways to formation of the transcription-preinitiation complex at the AT-rich TATA box of a template DNA molecule (see Figure 11-27). In both cases, multiprotein complexes assemble cooperatively through a network of specific protein–nucleic acid and protein–protein interactions.

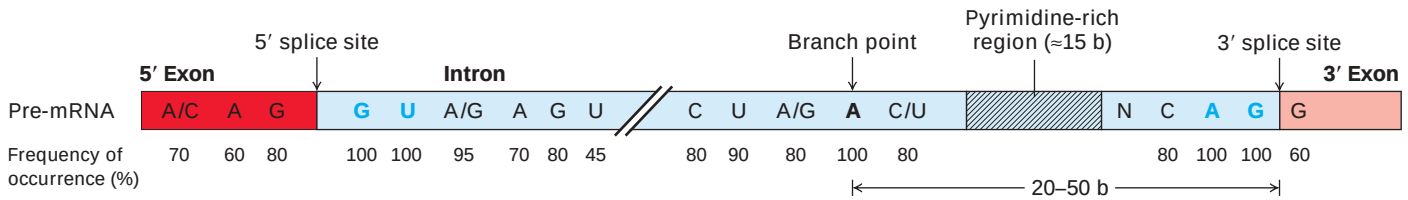
Following cleavage at the poly(A) site, polyadenylation proceeds in two phases. Addition of the first 12 or so A residues occurs slowly, followed by rapid addition of up to 200–250 more A residues. The rapid phase requires the binding of multiple copies of a *poly(A)-binding protein* containing the RRM motif. This protein is designated **PABPII** to distinguish it from the poly(A)-binding protein present in the cytoplasm. PABPII binds to the short A tail initially added by PAP, stimulating polymerization of additional A residues by PAP (see Figure 12-4). PABPII is also responsible for signaling poly(A) polymerase to terminate polymerization when the poly(A) tail reaches a length of 200–250 residues, although the mechanism for controlling the length of the tail is not yet understood.

Splicing Occurs at Short, Conserved Sequences in Pre-mRNAs via Two Transesterification Reactions

During formation of a mature, functional mRNA, the **introns** are removed and **exons** are spliced together. For short transcription units, **RNA splicing** usually follows cleavage and polyadenylation of the 3' end of the primary transcript, as depicted in Figure 12-2. However, for long transcription units



◀ **EXPERIMENTAL FIGURE 12-5 RNA-DNA hybridization studies show that introns are spliced out during pre-mRNA processing.** Electron microscopy of an RNA-DNA hybrid between adenovirus DNA and the mRNA encoding hexon, a major viral protein, reveals DNA segments (introns) that are absent from the hexon mRNA. (a) Diagram of the *EcoRI* A fragment of adenovirus DNA, which extends from the left end of the genome to just before the end of the final exon of the hexon gene. The gene consists of three short exons and one long (≈ 3.5 kb) exon separated by three introns of ≈ 1 , 2.5, and 9 kb. (b) Electron micrograph (*left*) and schematic drawing (*right*) of hybrid between an *EcoRI* A fragment and hexon mRNA. The loops marked A, B, and C correspond to the introns indicated in (a). Since these intron sequences in the viral genomic DNA are not present in mature hexon mRNA, they loop out between the exon sequences that hybridize to their complementary sequences in the mRNA. [Micrograph from S. M. Berget et al., 1977, *Proc. Nat'l. Acad. Sci. USA* **74**:3171; courtesy of P. A. Sharp.]



▲ **FIGURE 12-6 Consensus sequences around 5' and 3' splice sites in vertebrate pre-mRNAs.** The only nearly invariant bases are the 5' GU and the 3' AG of the intron (blue), although the flanking bases indicated are found at frequencies higher than expected based on a random distribution. A pyrimidine-rich region (hatch marked) near the 3' end of the intron is found in most

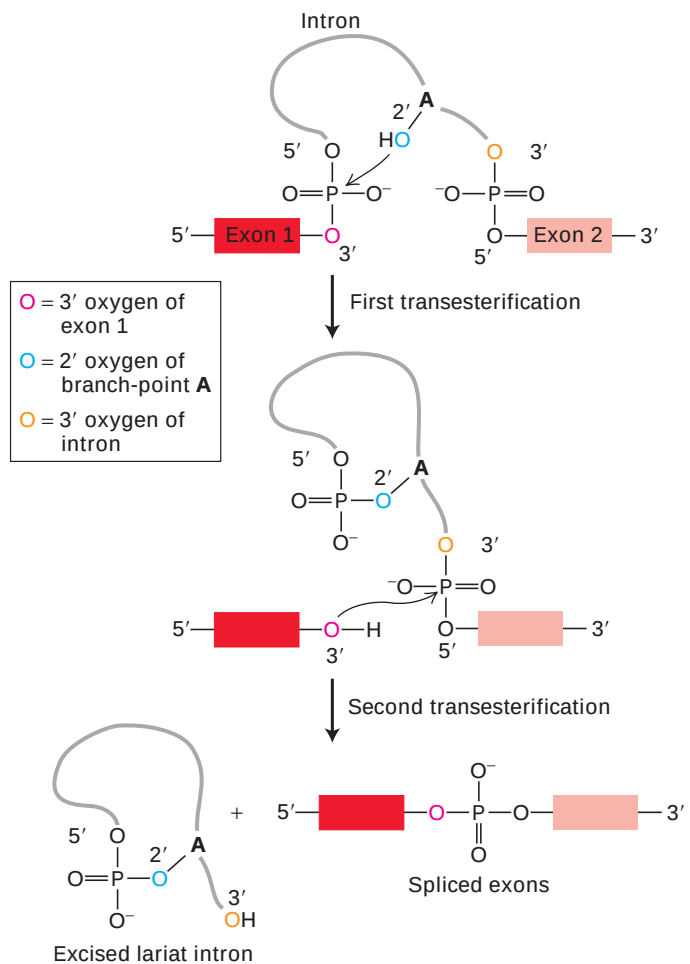
cases. The branch-point adenosine, also invariant, usually is 20–50 bases from the 3' splice site. The central region of the intron, which may range from 40 bases to 50 kilobases in length, generally is unnecessary for splicing to occur. [See R. A. Padgett et al., 1986, *Ann. Rev. Biochem.* **55**:1119, and E. B. Keller and W. A. Noon, 1984, *Proc. Nat'l. Acad. Sci. USA* **81**:7417.]

containing multiple exons, splicing of exons in the nascent RNA usually begins before transcription of the gene is complete.

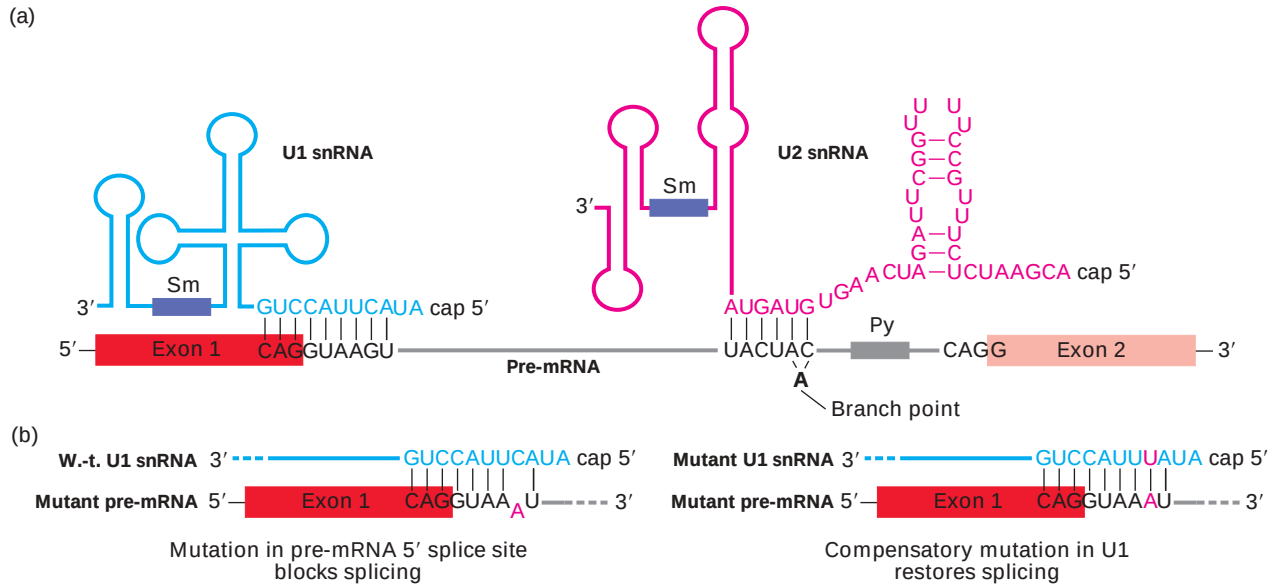
Early evidence that introns are removed during splicing came from electron microscopy of RNA-DNA hybrids between adenovirus DNA and the mRNA encoding hexon, a major virion capsid protein (Figure 12-5). Other studies revealed nuclear viral RNAs that were colinear with the viral DNA (primary transcripts) and RNAs with one or two of the introns removed (processing intermediates). These results, together with the findings that the 5' cap and 3' poly(A) tail at each end of long mRNA precursors are retained in shorter mature cytoplasmic mRNAs, led to the realization that introns are removed from primary transcripts as exons are spliced together.

The location of *splice sites*—that is, exon-intron junctions—in a pre-mRNA can be determined by comparing the sequence of genomic DNA with that of the cDNA prepared from the corresponding mRNA. Sequences that are present in the genomic DNA but absent from the cDNA represent introns and indicate the positions of splice sites. Such analysis of a large number of different mRNAs revealed moderately conserved, short consensus sequences at the splice sites flanking introns in eukaryotic pre-mRNAs; in higher organisms, a pyrimidine-rich region just upstream of the 3' splice site also is common (Figure 12-6). Studies with deletion mutants have shown that much of the center portion of introns can be removed without affecting splicing; generally only 30–40 nucleotides at each end of an intron are necessary for splicing to occur at normal rates.

Analysis of the intermediates formed during splicing of pre-mRNAs in vitro led to the discovery that splicing of exons proceeds via two sequential *transesterification reactions* (Figure 12-7). Introns are removed as a lariat-like structure in which the 5' G of the intron is joined in an unusual 2',5'-phosphodiester bond to an adenosine near the 3' end of the intron. This A residue is called the *branch point* because it forms an RNA branch in the lariat structure. In each transesterification reaction, one phosphoester bond is exchanged for another. Since the number of phosphoester bonds in the molecule is not changed in either reaction, no energy is consumed. The net result of these two reactions is that two exons are ligated and the intervening intron is released as a branched lariat structure.



▲ **FIGURE 12-7 Two transesterification reactions that result in splicing of exons in pre-mRNA.** In the first reaction, the ester bond between the 5' phosphorus of the intron and the 3' oxygen (dark red) of exon 1 is exchanged for an ester bond with the 2' oxygen (blue) of the branch-site A residue. In the second reaction, the ester bond between the 5' phosphorus of exon 2 and the 3' oxygen (light red) of the intron is exchanged for an ester bond with the 3' oxygen of exon 1, releasing the intron as a lariat structure and joining the two exons. Arrows show where the activated hydroxyl oxygens react with phosphorus atoms.



▲ **FIGURE 12-8 Base pairing between pre-mRNA, U1 snRNA, and U2 snRNA early in the splicing process.** (a) In this diagram, secondary structures in the snRNAs that are not altered during splicing are depicted schematically. The yeast branch-point sequence is shown here. Note that U2 snRNA base-pairs with a sequence that includes the branch-point A, although this residue is not base-paired. The purple rectangles represent sequences that bind snRNP proteins recognized by anti-Sm antibodies. For unknown reasons, antisera from patients with the autoimmune disease systemic lupus erythematosus (SLE) contain these antibodies. Such antisera have been useful in

characterizing components of the splicing reaction. (b) Only the 5' ends of U1 snRNAs and 5' splice sites in pre-mRNAs are shown. (Left) A mutation (A) in a pre-mRNA splice site that interferes with base-pairing to the 5' end of U1 snRNA blocks splicing. Expression of a U1 snRNA with a compensating mutation (U) that restores base-pairing also restores splicing of the mutant pre-mRNA. [Part (a) adapted from M. J. Moore et al., 1993, in R. Gesteland and J. Atkins, eds., *The RNA World*, Cold Spring Harbor Press, pp. 303–357; part (b) see Y. Zhuang and A. M. Weiner, 1986, *Cell* **46**:827.]

snRNAs Base-Pair with Pre-mRNA and with One Another During Splicing

Five U-rich **small nuclear RNAs (snRNAs)**, designated U1, U2, U4, U5, and U6, participate in pre-mRNA splicing. Ranging in length from 107 to 210 nucleotides, these snRNAs are associated with 6 to 10 proteins in *small nuclear ribonucleoprotein particles (snRNPs)* in the nucleus of eukaryotic cells.

Definitive evidence for the role of U1 snRNA in splicing came from experiments which indicated that base pairing between the 5' splice site of a pre-mRNA and the 5' region of U1 snRNA is required for RNA splicing (Figure 12-8a). In vitro experiments showed that a synthetic oligonucleotide that hybridizes with the 5'-end region of U1 snRNA blocks RNA splicing. In vivo experiments showed that mutations in the pre-mRNA 5' splice site that disrupt the base pairing block RNA splicing, but that splicing can be restored by expression of a mutant U1 snRNA with a compensating mutation that restores base pairing to the mutant pre-mRNA 5' splice site (Figure 12-8b). Involvement of U2 snRNA in splicing initially was suspected when it was found to have an internal sequence that is largely complementary to the consensus sequence flanking the branch

point in pre-mRNAs (see Figure 12-6). Compensating mutation experiments, similar to those conducted with U1 snRNA and 5' splice sites, demonstrated that base pairing between U2 snRNA and the branch-point sequence in pre-mRNA also is critical to splicing.

Figure 12-8a illustrates the general structures of the U1 and U2 snRNAs and how they base-pair with pre-mRNA during splicing. Significantly, the branch-point A itself, which is not base-paired to U2 snRNA, “bulges out,” allowing its 2' hydroxyl to participate in the first transesterification reaction of RNA splicing (see Figure 12-7).

Similar studies with other snRNAs demonstrated that base pairing between them also occurs during splicing. Moreover, rearrangements in these RNA-RNA interactions are critical in the splicing pathway, as we describe next.

Spliceosomes, Assembled from snRNPs and a Pre-mRNA, Carry Out Splicing

According to the current model of pre-mRNA splicing, the five splicing snRNPs are thought to assemble on the pre-mRNA, forming a large ribonucleoprotein complex called a **spliceosome** (Figure 12-9). Assembly of a spliceosome begins with the base pairing of the snRNAs of the U1 and U2

snRNPs to the pre-mRNA (see Figure 12-8). Extensive base pairing between the snRNAs in the U4 and U6 snRNPs forms a complex that associates with U5 snRNP. The U4/U6/U5 complex then associates with the previously formed U1/U2/pre-mRNA complex to yield a spliceosome.

After formation of the spliceosome, extensive rearrangements in the pairing of snRNAs and the pre-mRNA lead to release of the U1 and U4 snRNPs. The catalytically active rearranged spliceosome then mediates the first transesterification reaction that forms the 2',5'-phosphodiester bond between the 2' hydroxyl on the branch point A and the phos-

phate at the 5' end of the intron. Following another rearrangement of the snRNPs, the second transesterification reaction ligates the two exons in a standard 3',5'-phosphodiester bond, releasing the intron as a lariat structure associated with the snRNPs. This final intron-snRNP complex rapidly dissociates, and the individual snRNPs released can participate in a new cycle of splicing. The excised intron is then rapidly degraded by a debranching enzyme and other nuclear RNases discussed later.

A spliceosome is roughly the size of a small ribosomal subunit, and at least 70 proteins are thought to participate in RNA splicing, making this process comparable in complexity to initiation of transcription and protein synthesis. Some of these splicing factors are associated with snRNPs, but others are not. For instance, the 65-kD subunit of the U2-associated factor (U2AF) binds to the pyrimidine-rich region near the 3' end of introns and to the U2 snRNP. The 35-kD subunit of U2AF binds to the AG dinucleotide at the 3' end of the intron and also interacts with the larger U2AF subunit bound nearby. These two U2AF subunits act together to help specify the 3' splice site by promoting interaction of U2 snRNP with the branch point (see Figure 12-8). Some splicing factors also exhibit sequence homologies to known RNA helicases; these are probably necessary for the base-pairing rearrangements that occur in snRNAs during the spliceosomal splicing cycle.

Some pre-mRNAs contain introns whose splice sites do not conform to the standard consensus sequence. This class of introns begins with AU and ends with AC rather than following the usual "GU-AG rule" (see Figure 12-6). Splicing of this special class of introns appears to occur via a splicing cycle analogous to that shown in Figure 12-9, except that four novel, low-abundance snRNPs, together with the standard U5 snRNP, are involved.

Nearly all functional mRNAs in vertebrate, insect, and plant cells are derived from a single molecule of the corresponding pre-mRNA by removal of internal introns and splicing of exons. However, in two types of protozoans—trypanosomes and euglenoids—mRNAs are constructed by

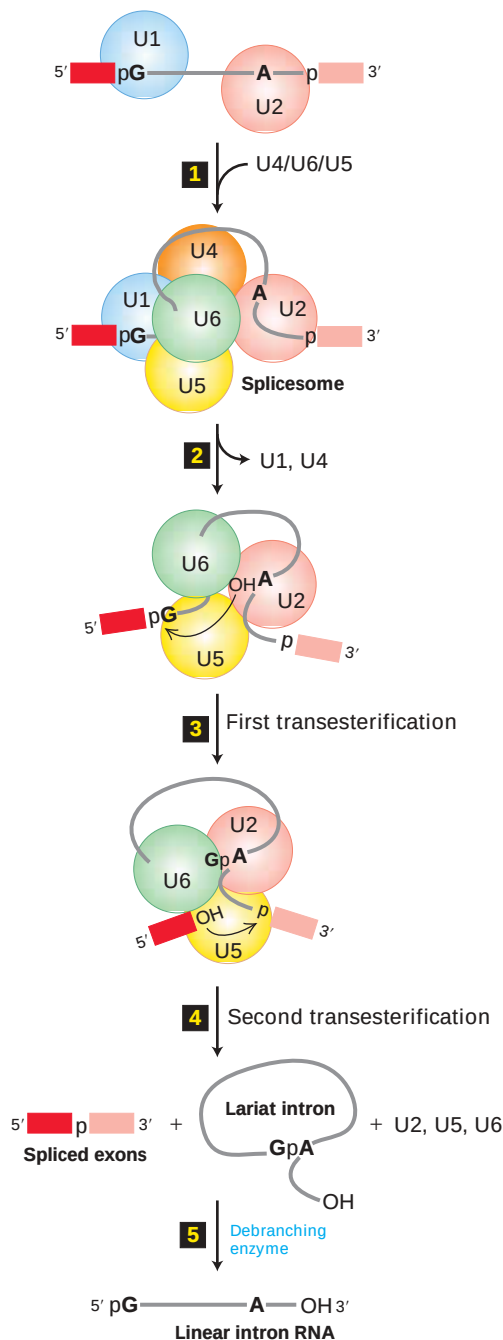
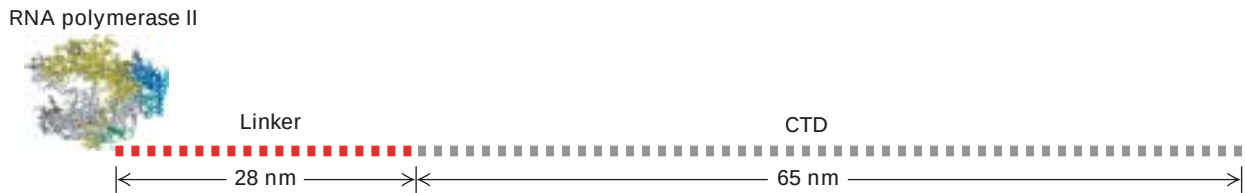


FIGURE 12-9 Model of spliceosome-mediated splicing of pre-mRNA. Step **1**: After U1 and U2 snRNPs associate with the pre-mRNA, via base-pairing interactions shown in Figure 12-8, a trimeric snRNP complex of U4, U5, and U6 joins the initial complex to form the spliceosome. Step **2**: Rearrangements of base-pairing interactions between snRNAs converts the spliceosome into a catalytically active conformation and destabilizes the U1 and U4 snRNPs, which are released. Step **3**: The catalytic core, thought to be formed by U6 and U2, then catalyzes the first transesterification reaction, forming the intermediate containing a 2',5'-phosphodiester bond shown in Figure 12-7. Step **4**: Following further rearrangements between the snRNPs, the second transesterification reaction joins the two exons by a standard 3',5'-phosphodiester bond and releases the intron as a lariat structure and the remaining snRNPs. Step **5**: The excised lariat intron is converted into a linear RNA by a debranching enzyme. [Adapted from T. Villa et al., 2002, *Cell* **109**:149.]



▲ **FIGURE 12-10 Schematic diagram of RNA polymerase II with the CTD extended.** The length of the fully extended yeast RNA polymerase II carboxyl-terminal domain (CTD) and the linker region that connects it to the polymerase is shown relative to the globular domain of the polymerase. The CTD of mammalian

RNA polymerase II is twice as long. In its extended form, the CTD can associate with multiple RNA-processing factors simultaneously. [From P. Cramer, D. A. Bushnell, and R. D. Kornberg, 2001, *Science* **292**:1863.]

splicing together separate RNA molecules. This process, referred to as *trans-splicing*, is also used in the synthesis of 10–15 percent of the mRNAs in the nematode (round worm) *Caenorhabditis elegans*, an important model organism for studying embryonic development. Trans-splicing is carried out by snRNPs by a process similar to the splicing of exons in a single pre-mRNA.

Chain Elongation by RNA Polymerase II Is Coupled to the Presence of RNA-Processing Factors

The carboxyl-terminal domain (CTD) of RNA polymerase II is composed of multiple repeats of a seven-residue (heptapeptide) sequence. When fully extended, the CTD domain in the yeast enzyme is about 65 nm long (Figure 12-10); the CTD in human polymerase is about twice as long. The remarkable length of the CTD apparently allows multiple proteins to associate simultaneously with a single RNA polymerase II molecule. For instance, as mentioned earlier, the enzyme that adds the 5' cap to nascent transcripts associates with the phosphorylated CTD shortly after transcription initiation. In addition, RNA splicing and polyadenylation factors have been found to associate with the phosphorylated CTD. As a consequence, these processing factors are present at high local concentrations when splice sites and poly(A) signals are transcribed by the polymerase, enhancing the rate and specificity of RNA processing.

Recent results indicate that the association of RNA splicing factors with the phosphorylated CTD also stimulates transcription elongation. Thus chain elongation is coupled to the binding of RNA-processing factors to the phosphorylated CTD. This mechanism may ensure that a pre-mRNA is not synthesized unless the machinery for processing it is properly positioned.

SR Proteins Contribute to Exon Definition in Long Pre-mRNAs

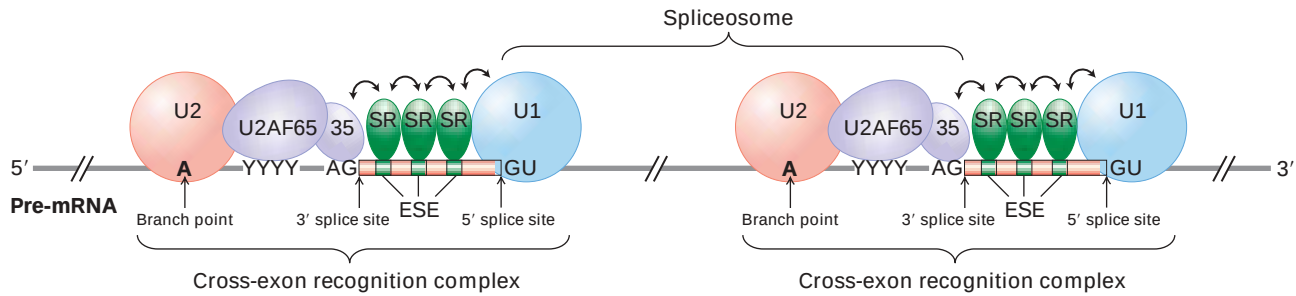
The average length of an exon in the human genome is ≈ 150 bases, whereas the average length of an intron is much longer (≈ 3500 bases). The longest introns contain upwards of 500 kb! Because the sequences of 5' and 3' splice sites and branch points are so degenerate, multiple copies are likely to occur

randomly in long introns. Consequently, additional sequence information is required to define the exons that should be spliced together in higher organisms with long introns.

The more sophisticated mechanism of exon recognition required in higher organisms entails a family of RNA-binding proteins, the *SR proteins*, that interact with sequences within exons called *exonic splicing enhancers*. SR proteins are characterized by having one or more RRM RNA-binding domains and several protein-protein interaction domains rich in serine and arginine residues. When bound to exonic splicing enhancers, SR proteins mediate the cooperative binding of U1 snRNP to a true 5' splice site and U2 snRNP to a branch point through a network of protein-protein interactions that span across an exon (Figure 12-11). The complex of SR proteins, snRNPs, and other splicing factors (e.g., U2AF) that assemble across an exon, which has been called a **cross-exon recognition complex**, permits precise specification of exons in long pre-mRNAs.



In the transcription units of higher organisms with long introns, exons not only encode the amino acid sequences of different portions of a protein but also contain binding sites for SR proteins. Mutations that interfere with the binding of an SR protein to an exonic splicing enhancer, even if they do not change the encoded amino acid sequence, would prevent formation of the cross-exon recognition complex. As a result, the affected exon is “skipped” during splicing and not included in the final processed mRNA. The truncated mRNA produced in this case is either degraded or translated into a mutant, abnormally functioning protein. Recent studies have implicated this type of mutation in human genetic diseases. For example, *spinal muscle atrophy* is one of the most common genetic causes of childhood mortality. This disease results from mutations in a region of the genome containing two closely related genes, *SMN1* and *SMN2*, that arose by gene duplication. *SMN2* encodes a protein identical with *SMN1*, but it is expressed at much lower level because a silent mutation in one exon interferes with the binding of an SR protein, leading to exon skipping in most of the *SMN2* mRNAs. The homologous *SMN* gene in the mouse, where there is only a single copy, is essential for cell viability. Spinal muscle atrophy in humans results from homozygous mutations that inactivate *SMN1*. The low level of protein translated from the small



▲ FIGURE 12-11 Exon recognition through cooperative binding of SR proteins and splicing factors to pre-mRNA.

The correct 5' GU and 3' AG splice sites are recognized by splicing factors on the basis of their proximity to exons. The exons contain exonic splicing enhancers (ESEs) that are binding sites for SR proteins. When bound to ESEs, the SR proteins interact with one another and promote the cooperative binding of the U1 snRNP to the 5' splice site of the downstream intron, the U2 snRNP to the branch point of the upstream intron, the 65- and 35-kD subunits of U2AF to the pyrimidine-rich region and AG 3' splice site of the upstream intron, and other splicing factors (not shown). The resulting RNA-protein cross-exon

recognition complex spans an exon and activates the correct splice sites for RNA splicing. Note that the U1 and U2 snRNPs in this unit are not part of the same spliceosome. The U2 snRNP on the right forms a spliceosome with the U1 snRNP bound to the 5'-end of the same intron. The U1 snRNP shown on the right forms a spliceosome with the U2 snRNP bound to the branch point of the downstream intron (not shown), and the U2 snRNP on the left forms a spliceosome with a U1 snRNP bound to the 5' splice site of the upstream intron (not shown). Double-headed arrows indicate protein-protein interactions. [Adapted from T. Maniatis, 2002, *Nature* **418**:236; see also S. M. Berget, 1995, *J. Biol. Chem.* **270**:2411]

fraction of *SMN2* mRNAs that are correctly spliced is sufficient to maintain cell viability during embryogenesis and fetal development, but it is not sufficient to maintain viability of spinal cord motor neurons in childhood, resulting in their death and the associated disease.

Approximately 15 percent of the single-base-pair mutations that cause human genetic diseases interfere with proper exon definition. Some of these mutations occur in 5' or 3' splice sites, often resulting in the use of nearby alternative "cryptic" splice sites present in the normal gene sequence. In the absence of the normal splice site, the cross-exon recognition complex recognizes these alternative sites. Other mutations that cause abnormal splicing result in a new consensus splice site sequence that becomes recognized in place of the normal splice site. Finally, some mutations can interfere with the binding of specific hnRNP proteins to pre-mRNAs that enhance or repress splicing at normal splice sites, as in the case of the *SMN2* gene. ■

Self-Splicing Group II Introns Provide Clues to the Evolution of snRNAs

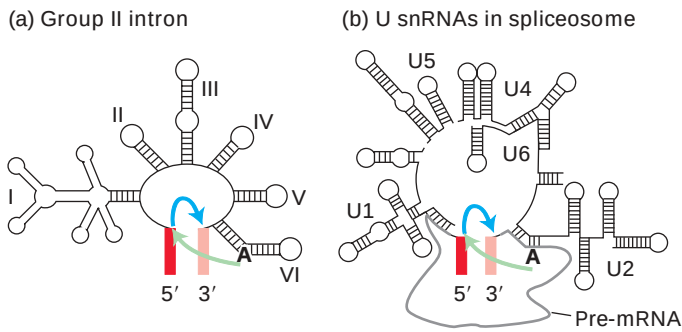
Under certain nonphysiological *in vitro* conditions, pure preparations of some RNA transcripts slowly splice out introns in the absence of any protein. This observation led to recognition that some introns are *self-splicing*. Two types of self-splicing introns have been discovered: *group I introns*, present in nuclear rRNA genes of protozoans, and *group II introns*, present in protein-coding genes and some rRNA and tRNA genes in mitochondria and chloroplasts of plants and fungi. Discovery of the catalytic activity of self-splicing introns revolutionized concepts about the functions of RNA. As discussed in Chapter 4, RNA is now thought to catalyze

peptide-bond formation during protein synthesis in ribosomes. Here we discuss the probable role of group II introns, now found only in mitochondrial and chloroplast DNA, in the evolution of snRNAs; the functioning of group I introns is considered in the later section on rRNA processing.

Even though their precise sequences are not highly conserved, all group II introns fold into a conserved, complex secondary structure containing numerous stem-loops (Figure 12-12a). Self-splicing by a group II intron occurs via two transesterification reactions, involving intermediates and products analogous to those found in nuclear pre-mRNA splicing. The mechanistic similarities between group II intron self-splicing and spliceosomal splicing led to the hypothesis that snRNAs function analogously to the stem-loops in the secondary structure of group II introns. According to this hypothesis, snRNAs interact with 5' and 3' splice sites of pre-mRNAs and with each other to produce a three-dimensional RNA structure functionally analogous to that of group II self-splicing introns (Figure 12-12b).

An extension of this hypothesis is that introns in ancient pre-mRNAs evolved from group II self-splicing introns through the progressive loss of internal RNA structures, which concurrently evolved into trans-acting snRNAs that perform the same functions. Support for this type of evolutionary model comes from experiments with group II intron mutants in which domain V and part of domain I are deleted. RNA transcripts containing such mutant introns are defective in self-splicing, but when RNA molecules equivalent to the deleted regions are added to the *in vitro* reaction, self-splicing occurs. This finding demonstrates that these domains in group II introns can be trans-acting, like snRNAs.

The similarity in the mechanisms of group II intron self-splicing and spliceosomal splicing of pre-mRNAs also



▲ **FIGURE 12-12 Schematic diagrams comparing the secondary structures of group II self-splicing introns (a) and U snRNAs present in the spliceosome (b).** The first transesterification reaction is indicated by green arrows; the second reaction, by blue arrows. The branch-point A is boldfaced. The similarity in these structures suggests that the spliceosomal snRNAs evolved from group II introns, with the trans-acting snRNAs being functionally analogous to the corresponding domains in group II introns. The colored bars flanking the introns in (a) and (b) represent exons [Adapted from P. A. Sharp, 1991, *Science* **254**:663.]

suggests that the splicing reaction is catalyzed by the snRNA, not the protein, components of spliceosomes. Although group II introns can self-splice in vitro at elevated temperatures and Mg^{2+} concentrations, under in vivo conditions proteins called *maturases*, which bind to group II intron RNA, are required for rapid splicing. Maturases are thought to stabilize the precise three-dimensional interactions of the intron RNA required to catalyze the two splicing transesterification reactions. By analogy, snRNP proteins in spliceosomes are

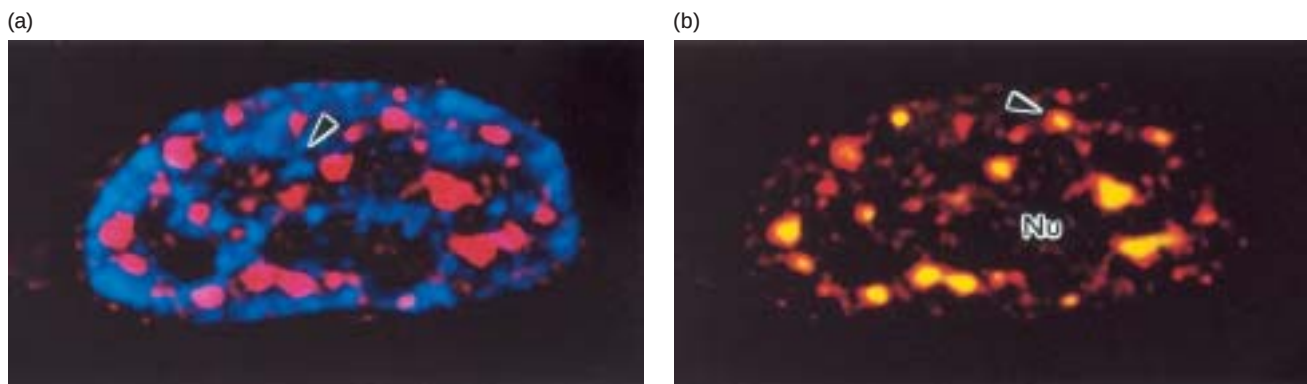
thought to stabilize the precise geometry of snRNAs and intron nucleotides required to catalyze pre-mRNA splicing.

The evolution of snRNAs may have been an important step in the rapid evolution of higher eukaryotes. As internal intron sequences were lost and their functions in RNA splicing supplanted by trans-acting snRNAs, the remaining intron sequences would be free to diverge. This in turn likely facilitated the evolution of new genes through **exon shuffling** since there are few constraints on the sequence of new introns generated in the process (see Figure 10-17). It also permitted the increase in protein diversity that results from alternative RNA splicing and an additional level of gene control resulting from regulated RNA splicing.

Most Transcription and RNA Processing Occur in a Limited Number of Domains in Mammalian Cell Nuclei

Several kinds of studies suggest that transcription and RNA processing occur largely within discrete foci within the nucleus of eukaryotic cells. For instance, digital imaging microscopy of human fibroblasts reveals that most of the nuclear polyadenylated RNA (i.e., unspliced and partially spliced pre-mRNA and nuclear mRNA) is localized in about 100 foci (Figure 12-13). An SR protein involved in exon definition (SC-35) is localized to the center of these same loci.

These data imply that the nucleus is not simply an unorganized container for chromatin and the proteins that carry out chromosome replication, transcription, and RNA processing. Rather, there is a highly organized underlying nuclear substructure. A poorly understood fibrillar network comprising multiple proteins called the *nuclear matrix* can be



▲ **EXPERIMENTAL FIGURE 12-13 Staining shows that polyadenylated RNA and RNA splicing factors are sequestered in discrete areas of the mammalian fibroblast nucleus.** Digital imaging microscopy was used to reconstruct a 1- μ m-thick section of a stained human fibroblast nucleus. (a) Section stained with red rhodamine-labeled poly(dT) to detect polyadenylated RNA (red) and with DAPI to detect DNA (blue). Polyadenylated RNA is localized to a limited number of discrete foci (speckles) between

regions of chromatin, although not all regions containing low levels of DNA contained detectable polyadenylated RNA (arrow). (b) The same section shown in (a) stained to detect polyadenylated RNA (red) and SR protein SC-35, which was visualized with a green fluorescein-labeled monoclonal antibody. Regions where the stains overlap appear yellow. SC-35 is present in the center of many foci (arrow). Nu = nucleolus. [From K. C. Carter et al., 1993, *Science* **259**:1330.]

observed in the nucleus after digestion of most DNA and extraction with high salt. This network of fibrous proteins probably contributes to nuclear organization just as cytoskeletal fibers contribute to the organization of the cytoplasm. Much remains to be learned about the molecular mechanisms underlying this organization and its consequences for nuclear processes.

Nuclear Exonucleases Degrade RNA That Is Processed out of Pre-mRNAs

Because the human genome contains long introns, only ≈ 5 percent of the nucleotides that are polymerized by RNA polymerase II during transcription are retained in the mature, processed mRNA. The introns that are spliced out and the region downstream from the cleavage and polyadenylation site are degraded by nuclear exonucleases that hydrolyze one base at a time from either the 5' or 3' end of an RNA strand.

As mentioned earlier, the 2',5'-phosphodiester bond in excised introns is hydrolyzed by a debranching enzyme, yielding a linear molecule with unprotected ends that can be attacked by exonucleases (see Figure 12-9). The predominant nuclear decay pathway is 3' $\rightarrow$ 5' hydrolysis by 11 exonucleases that associate with one another in a large protein complex called the **exosome**. Other proteins in the complex include RNA helicases that disrupt base pairing and RNA-protein interactions that would otherwise impede the exonucleases. Exosomes also function in the cytoplasm as discussed later. In addition the exosome appears to degrade pre-mRNAs that have not been properly spliced or polyadenylated. It is not yet clear how the exosome recognizes improperly processed pre-mRNAs.

To avoid being degraded by nuclear exonucleases, nascent transcripts, pre-mRNA processing intermediates, and mature mRNAs in the nucleus must have their ends protected. The 5' end of a nascent transcript is protected by addition of the 5' cap structure as soon as the 5' end emerges from the polymerase. Further protection is provided by interaction of the cap with a *nuclear cap-binding complex*, which also functions in export of mRNA to the cytoplasm. The 3' end of a nascent transcript lies within the RNA polymerase and thus is inaccessible to exonucleases (see Figure 4-10). As discussed previously, the free 3' end generated by cleavage of a pre-mRNA downstream from the poly(A) signal is rapidly polyadenylated by the poly(A) polymerase associated with the other 3' processing factors, and the resulting poly(A) tail is bound by PABPII (see Figure 12-4). This tight coupling of cleavage and polyadenylation protects the 3' end from exonuclease attack.

- Shortly after transcription initiation, a capping enzyme associated with the phosphorylated CTD of RNA polymerase II adds the 5' cap to the nascent transcript.

- In most protein-coding genes, a conserved AAUAAA poly(A) signal lies slightly upstream from a poly(A) site where cleavage and polyadenylation occur. A GU- or U-rich sequence downstream from the poly(A) site contributes to the efficiency of cleavage and polyadenylation.

- A multiprotein complex that includes poly(A) polymerase (PAP) carries out the cleavage and polyadenylation of a pre-mRNA. A nuclear poly(A)-binding protein, PABPII, stimulates addition of A residues by PAP and stops addition once the poly(A) tail reaches 200–250 residues (see Figure 12-4).

- Five different snRNPs interact via base pairing with one another and with pre-mRNA to form the spliceosome (see Figure 12-9). This very large ribonucleoprotein complex catalyzes two transesterification reactions that join two exons and remove the intron as a lariat structure, which is subsequently degraded (see Figure 12-7).

- The association of pre-mRNA processing factors with the CTD of RNA polymerase II stimulates chain elongation. This coupling ensures that a pre-mRNA is not synthesized until the factors required for its processing are in place to interact with splicing and cleavage and polyadenylation signals in the pre-mRNA as they emerge from the polymerase.

- SR proteins that bind to exonic splicing enhancer sequences in exons are critical in defining exons in the large pre-mRNAs of higher organisms. A network of interactions between SR proteins, snRNPs, and splicing factors forms a cross-exon recognition complex that specifies correct splice sites (see Figure 12-11).

- The snRNAs in the spliceosome are thought to have an overall tertiary structure similar to that of group II self-splicing introns.

- For long transcription units in higher organisms, splicing of exons usually begins as the pre-mRNA is still being formed. Cleavage and polyadenylation to form the 3' end of the mRNA occur after the poly(A) site is transcribed.

- Excised introns are degraded primarily by exosomes, multiprotein complexes that contain eleven 3' $\rightarrow$ 5' exonucleases as well as RNA helicases. Exosomes also degrade improperly processed pre-mRNAs.

KEY CONCEPTS OF SECTION 12.1

Processing of Eukaryotic Pre-mRNA

- In the nucleus of eukaryotic cells, pre-mRNAs are associated with hnRNP proteins and processed by 5' capping, 3' cleavage and polyadenylation, and splicing before being transported to the cytoplasm (see Figure 12-2).

12.2 Regulation of Pre-mRNA Processing

Now that we've seen how pre-mRNAs are processed into mature, functional mRNAs, we consider how regulation of this process can contribute to gene control. Recall from Chapter 10 that higher eukaryotes contain both simple and complex

transcription units. The primary transcripts produced from the former contain one poly(A) site and exhibit only one pattern of RNA splicing, even if multiple introns are present; thus simple transcription units encode a single mRNA. In contrast, the primary transcripts produced from complex transcription units can be processed in alternative ways to yield different mRNAs that encode distinct proteins (see Figure 10-2).

Alternative Splicing Is the Primary Mechanism for Regulating mRNA Processing

The results from recent genomic studies suggest that the number of complex transcription units may be much greater than previously realized. Researchers estimate that up to ≈60 percent of all transcription units in the human genome are complex. This estimate comes from comparison of the genomic DNA sequence with expressed sequence tags (ESTs), which are randomly sequenced segments of cDNA (Chapter 9). Comparison of these cDNA sequences with genomic DNA sequence shows the locations of exons and introns in genomic DNA. The finding of alternative combinations of exons expressed from a region of genomic DNA reveals the existence of a complex transcription unit. Even though this estimate may be elevated somewhat by inclusion in the EST database of rare RNAs that result from mistakes in RNA processing, it is clear that complex transcription units account for a large portion of the protein-coding sequences in humans. This finding and the discovery of multiple examples of regulated RNA processing indicate that alternative RNA processing of complex pre-mRNAs produced from such transcription units is a significant gene-control mechanism in higher eukaryotes.

Although examples of cleavage at alternative poly(A) sites in pre-mRNAs are known, *alternative splicing* of different exons is the more common mechanism for expressing different proteins from one complex transcription unit. Such alternative processing pathways usually are regulated, often in a cell-type-specific manner; that is, one possible mRNA is expressed in one type of cell or tissue, while another is expressed in different cells or tissues. In Chapter 4, for example, we mentioned that fibroblasts produce one type of the extracellular protein fibronectin, whereas hepatocytes produce another type. Both fibronectin **isoforms** are encoded by the same transcription unit, which is spliced differently in the two cell types to yield two different mRNAs (see Figure 4-15). In other cases, alternative processing may occur in the same cell type in response to different developmental or environmental signals. First we discuss one of the best-understood examples of regulated RNA processing and then consider its consequences in development of the nervous system.

A Cascade of Regulated RNA Splicing Controls *Drosophila* Sexual Differentiation

One of the earliest examples of regulated alternative splicing of pre-mRNA came from studies of sexual differentiation

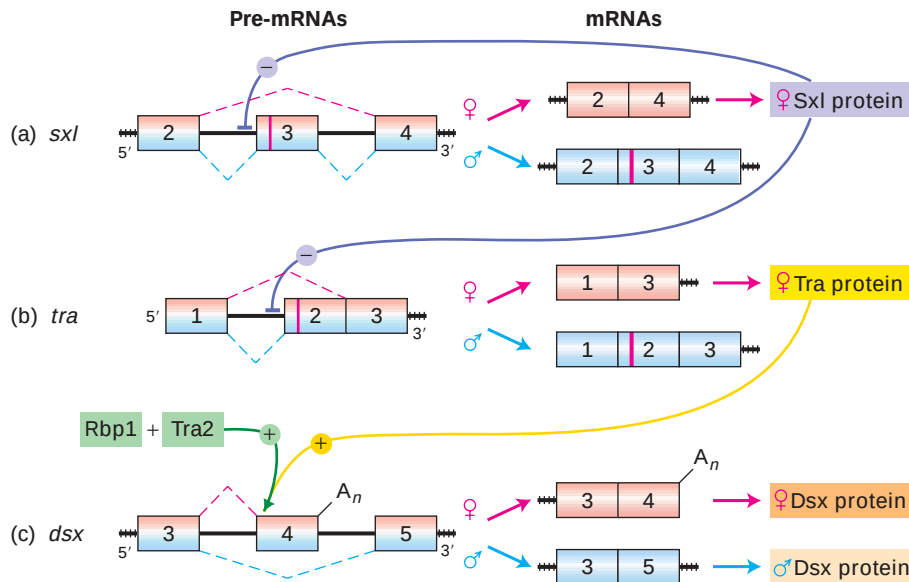
in *Drosophila*. Genes required for normal *Drosophila* sexual differentiation were first characterized by isolating *Drosophila* mutants defective in the process. When the proteins encoded by the wild-type genes were characterized biochemically, two of them were found to regulate a cascade of alternative RNA splicing in *Drosophila* embryos. More recent research has provided insight into how these proteins regulate RNA processing.

The Sxl protein, encoded by the *sex-lethal* gene, is the first protein to act in the cascade. Early in development, this gene is transcribed from a promoter that functions only in female embryos. Later in development, this female-specific promoter is shut off and another promoter for *sex-lethal* becomes active in both male and female embryos. However, in the absence of early Sxl protein, the *sex-lethal* pre-mRNA in male embryos is spliced to produce an mRNA that contains a stop codon early in the sequence. The net result is that male embryos produce no functional Sxl protein either early or later in development.

In contrast, the Sxl protein expressed in early female embryos directs splicing of the *sex-lethal* pre-mRNA so that a functional *sex-lethal* mRNA is produced (Figure 12-14a). Sxl accomplishes this by binding to a sequence in the pre-mRNA near the 3' end of the intron between exon 2 and exon 3, thereby blocking the proper association of U2AF and U2 snRNP. As a consequence, the U1 snRNP bound to the 3' end of exon 2 assembles into a spliceosome with U2 snRNP bound to the branch point at the 3' end of the intron between exons 3 and 4, leading to splicing of exon 2 to 4 and skipping of exon 3. The resulting female-specific *sex-lethal* mRNA is translated into functional Sxl protein, which reinforces its own expression in female embryos by continuing to cause skipping of exon 3. The absence of Sxl protein in male embryos allows the inclusion of exon 3 and, consequently, of the stop codon that prevents translation of functional Sxl protein.

Sxl protein also regulates alternative RNA splicing of the *transformer* gene pre-mRNA (Figure 12-14b). In male embryos, where no Sxl is expressed, exon 1 is spliced to exon 2, which contains a stop codon that prevents synthesis of a functional protein. In female embryos, however, binding of Sxl protein to the 3' end of the intron between exons 1 and 2 blocks binding of U2AF at this site. The interaction of Sxl with *transformer* pre-mRNA is mediated by two RRM domains in the protein (see Figure 12-3). When Sxl is bound, U2AF binds to a lower-affinity site farther 3' in the pre-mRNA; as a result exon 1 is spliced to this alternative 3' splice site, eliminating exon 2 with its stop codon. The resulting female-specific *transformer* mRNA, which contains additional constitutively spliced exons, is translated into functional Transformer (Tra) protein.

Finally, Tra protein regulates the alternative processing of pre-mRNA transcribed from the *double-sex* gene (Figure 12-14c). In female embryos, a complex of Tra and two constitutively expressed proteins, Rbp1 and Tra2, directs splicing of exon 3 to exon 4 and also promotes cleavage/



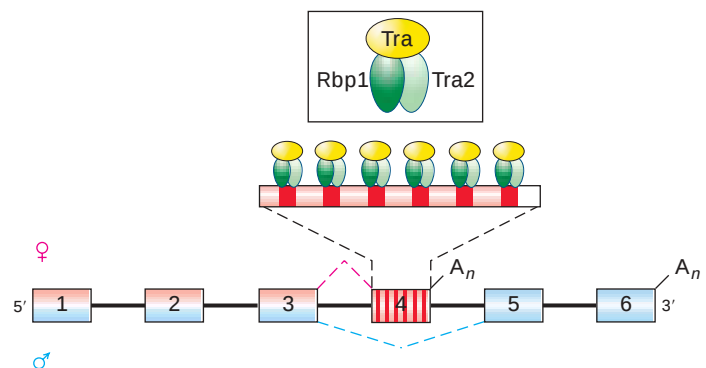
▲ **FIGURE 12-14 Cascade of regulated splicing that controls sex determination via expression of *sex-lethal* (*sxl*), *transformer* (*tra*), and *double-sex* (*dsx*) genes in *Drosophila* embryos.** For clarity, only the exons (boxes) and introns (black lines) where regulated splicing occurs are shown. Splicing is indicated by red dashed lines above (female) and blue dashed lines below (male) the pre-mRNAs. Vertical red lines in exons indicate in-frame stop codons, which prevent synthesis of functional protein. Only female embryos produce functional Sxl protein, which *represses* splicing between exons 2 and 3 in *sxl* pre-mRNA (a) and between exons 1 and 2 in *tra* pre-mRNA (b).

(c) In contrast, the cooperative binding of Tra protein and two SR proteins, Rbp1 and Tra2, *activates* splicing between exons 3 and 4 and cleavage/polyadenylation A_n at the 3' end of exon 4 in *dsx* pre-mRNA in female embryos. In male embryos, which lack functional Tra, the SR proteins do not bind to exon 4, and consequently exon 3 is spliced to exon 5. The distinct Dsx proteins produced in female and male embryos as the result of this cascade of regulated splicing repress transcription of genes required for sexual differentiation of the opposite sex. [Adapted from M. J. Moore et al., 1993, in R. Gesteland and J. Atkins, eds., *The RNA World*, Cold Spring Harbor Press, pp. 303–357.]

polyadenylation at the alternative poly(A) site at the 3' end of exon 4. In male embryos, which produce no Tra protein, exon 4 is skipped, so that exon 3 is spliced to exon 5. Exon 5 is constitutively spliced to exon 6, which is polyadenylated at its 3' end. As a result of the cascade of regulated RNA processing depicted in Figure 12-14, different Dsx proteins are expressed in male and female embryos. The male Dsx protein is a transcriptional repressor that inhibits the expression of genes required for female development. Conversely, the female Dsx protein represses transcription of genes required for male development.

Figure 12-15 illustrates how the Tra/Tra2/Rbp1 complex is thought to interact with *double-sex* pre-mRNA. Recent studies have shown that Rbp1 and Tra2 are SR proteins, similar to those discussed previously. They mediate the cooperative binding of the Tra/Tra2/Rbp1 complex to six exonic splicing enhancers in exon 4. The bound Tra2 and Rbp1 proteins then promote the binding of U2AF and U2 snRNP to the 3' end of the intron between exons 3 and 4, just as other SR proteins do for constitutively spliced exons (see Figure 12-11). The Tra/Tra2/Rbp1 complexes may also enhance binding of the cleavage/polyadenylation complex to the 3' end of exon 4.

► **FIGURE 12-15 Model of splicing activation by Tra protein and the SR proteins Rbp1 and Tra2.** In female *Drosophila* embryos, splicing of exons 3 and 4 in *dsx* pre-mRNA is activated by binding of Tra/Tra2/Rbp1 complexes to six sites in exon 4. Because Rbp1 and Tra2 cannot bind to the pre-mRNA in the absence of Tra, exon 4 is skipped in male embryos. See the text for discussion. A_n = polyadenylation. [Adapted from T. Maniatis and B. Tasic, 2002, *Nature* **418**:236.]



Splicing Repressors and Activators Control Splicing at Alternative Sites

As is evident from Figure 12-14, the *Drosophila* Sxl protein and Tra protein have opposite effects: Sxl prevents splicing, causing exons to be skipped, whereas Tra promotes splicing. The action of similar proteins may explain the cell-type-specific expression of fibronectin isoforms in humans. For instance, an Sxl-like splicing repressor expressed in hepatocytes might bind to splice sites for the EIIIA and EIIIB exons in the fibronectin pre-mRNA, causing them to be skipped during RNA splicing (see Figure 4-15). Alternatively, a Tra-like splicing activator expressed in fibroblasts might activate the splice sites associated with the fibronectin EIIIA and EIIIB exons, leading to inclusion of these exons in the mature mRNA. Experimental examination in some systems has revealed that inclusion of an exon in some cell types versus skipping of the same exon in other cell types results from the combined influence of several splicing repressors and enhancers.

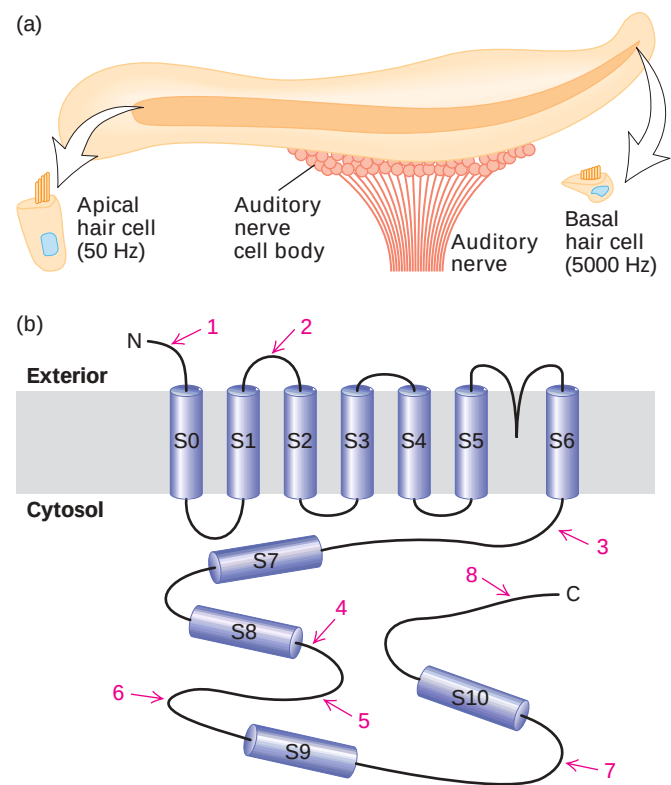
Alternative splicing of exons is especially common in the nervous system, generating multiple isoforms of many proteins required for neuronal development and function in both vertebrates and invertebrates. The primary transcripts from these genes often show quite complex splicing patterns that can generate several different mRNAs, with different spliced forms expressed in different anatomical locations within the central nervous system. We consider two remarkable examples that illustrate the critical role of this process in neural function.

Expression of *Slo* Isoforms in Vertebrate Hair Cells In the inner ear of vertebrates, individual “hair cells,” which are ciliated neurons, respond most strongly to a specific frequency of sound. Cells tuned to low frequency (≈ 50 Hz) are found at one end of the tubular cochlea that makes up the inner ear; cells responding to high frequency (≈ 5000 Hz) are found at the other end (Figure 12-16a). Cells in between respond to a gradient of frequencies between these extremes. One component in the tuning of hair cells in reptiles and birds is the opening of K^+ ion channels in response to increased intracellular Ca^{2+} concentrations. The Ca^{2+} concentration at which the channel opens determines the frequency with which the membrane potential oscillates, and hence the frequency to which the cell is tuned.

The gene encoding this channel (called *slo*, after the homologous *Drosophila* gene) is expressed as multiple, alternatively spliced mRNAs. The various Slo proteins encoded by these alternative mRNAs open at different Ca^{2+} concentrations. Hair cells with different response frequencies express different isoforms of the Slo channel protein depending on their position along the length of the cochlea. The sequence variation in the protein is very complex: there are at least eight regions in the mRNA where alternative exons are utilized, permitting the expression of 576 possible isoforms (Figure 12-16b). PCR analysis of *slo* mRNAs from individual hair cells has shown that each hair cell expresses a mixture of

different alternative *slo* mRNAs, with different forms predominating in different cells according to their position along the cochlea. This remarkable arrangement suggests that splicing of the *slo* pre-mRNA is regulated in response to extracellular signals that inform the cell of its position along the cochlea.

Recent results have shown that splicing at one of the alternative splice sites in the *slo* pre-mRNA in the rat is suppressed when a specific protein kinase is activated by depolarization of the neuron in response to synaptic activity from interacting neurons. This observation raises the possibility that a splicing repressor specific for this site may be activated when it is phosphorylated by this protein kinase,



▲ FIGURE 12-16 Role of alternative splicing of *slo* mRNA in the perception of sounds of different frequency. (a) The chicken cochlea, a 5-mm-long tube, contains an epithelium of auditory hair cells that are tuned to a gradient of vibrational frequencies from 50 Hz at the apical end (left) to 5000 Hz at the basal end (right). (b) The Slo protein contains seven transmembrane α helices (S0–S6), which associate to form the K^+ channel. The cytosolic domain, which includes four hydrophobic regions (S7–S10), regulates opening of the channel in response to Ca^{2+} . Isoforms of the Slo channel, encoded by alternatively spliced mRNAs produced from the same primary transcript, open at different Ca^{2+} concentrations and thus respond to different frequencies. Red numbers refer to regions where alternative splicing produces different amino acid sequences in the various Slo isoforms. [Adapted from K. P. Rosenblatt et al., 1997, *Neuron* 19:1061.]

whose activity in turn is regulated by synaptic activity. Since hnRNP and SR proteins are extensively modified by phosphorylation and other post-translational modifications, it seems likely that complex regulation of alternative RNA splicing through post-translational modifications of splicing factors plays a significant role in modulating neuron function.

Expression of *Dscam* Isoforms in *Drosophila* Retinal Neurons The most extreme example of regulated alternative RNA processing yet uncovered occurs in expression of the *Dscam* gene in *Drosophila*. Mutations in this gene interfere with the normal connections made by the axons of retinal neurons with neurons in a specific region of the brain during fly development. Analysis of the *Dscam* gene shows that it contains 95 alternatively spliced exons that could be spliced to generate over 38,000 possible isoforms! These results raise the possibility that the expression of different *Dscam* isoforms through regulated RNA splicing helps to specify the tens of thousands of different specific synaptic connections made between retinal and brain neurons. In other words, the correct wiring of neurons in the brain may depend on regulated RNA splicing.

RNA Editing Alters the Sequences of Pre-mRNAs

Sequencing of numerous cDNA clones and of the corresponding genomic DNAs from multiple organisms led in the mid-1980s to the unexpected discovery of a previously unrecognized type of pre-mRNA processing. In this type of processing, called **RNA editing**, the sequence of a pre-mRNA is altered; as a result, the sequence of the corresponding mature mRNA differs from the exons encoding it in genomic DNA.

RNA editing is widespread in the mitochondria of protozoans and plants and also in chloroplasts; in these organelles, more than half the sequence of some mRNAs is altered from the sequence of the corresponding primary tran-

scripts. In higher eukaryotes, RNA editing is much rarer, and thus far only single-base changes have been observed. Such minor editing, however, turns out to have important functional consequences in some cases.

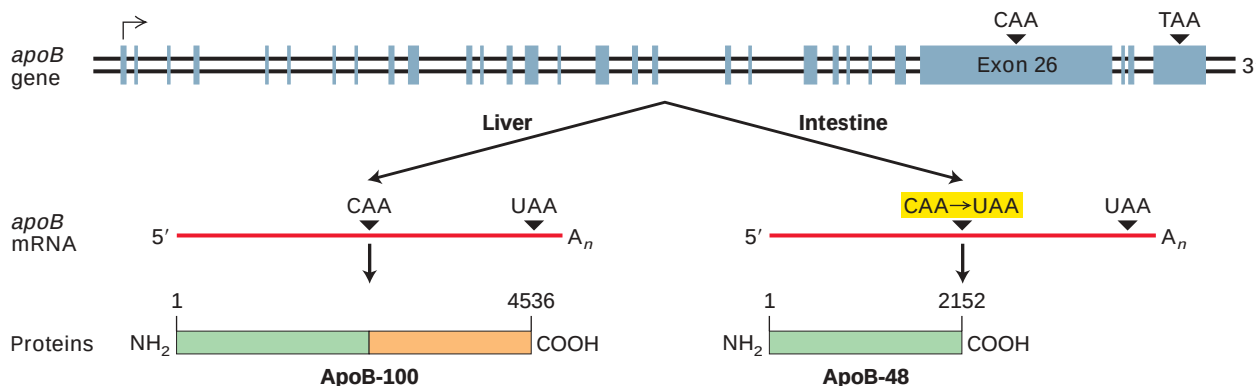
An important example of RNA editing in mammals involves the *apoB* gene, which encodes two alternative forms of the serum protein apolipoprotein B (apoB): apoB-100 expressed in hepatocytes and apoB-48 expressed in intestinal epithelial cells. The ≈240-kDa apoB-48 corresponds to the N-terminal region of the ≈500-kDa apoB-100. As we detail in Chapter 18, both apoB proteins are components of large lipoprotein complexes that transport lipids in the serum. However, only low-density lipoprotein (LDL) complexes, which contain apoB-100 on their surface, deliver cholesterol to body tissues by binding to the LDL receptor present on all cells.

The cell-type-specific expression of the two forms of apoB results from editing of *apoB* pre-mRNA so as to change the nucleotide at position 6666 in the sequence from a C to a U. This alteration, which occurs only in intestinal cells, converts a CAA codon for glutamine to a UAA stop codon, leading to synthesis of the shorter apoB-48 (Figure 12-17). Studies with the partially purified enzyme that performs the post-transcriptional deamination of C₆₆₆₆ to U shows that it can recognize and edit an RNA as short as 26 nucleotides with the sequence surrounding C₆₆₆₆ in the *apoB* primary transcript.

KEY CONCEPTS OF SECTION 12.2

Regulation of Pre-mRNA Processing

- Because of alternative splicing of primary transcripts and cleavage at different poly(A) sites, different mRNAs may be expressed from the same gene in different cell types or at different developmental stages (see Figure 12-14).
- Alternative splicing can be regulated by RNA-binding proteins that bind to specific sequences near regulated



▲ **FIGURE 12-17 RNA editing of *apoB* pre-mRNA.** The *apoB* mRNA produced in the liver has the same sequence as the exons in the primary transcript. This mRNA is translated into apoB-100, which has two functional domains: an N-terminal domain (green) that associates with lipids and a C-terminal domain (orange) that binds to LDL receptors on cell membranes.

In the *apoB* mRNA produced in the intestine, the CAA codon in exon 26 is edited to a UAA stop codon. As a result, intestinal cells produce apoB-48, which corresponds to the N-terminal domain of apoB-100. [Adapted from P. Hodges and J. Scott, 1992, *Trends Biochem. Sci.* **17**:77.]

splice sites. Splicing repressors may sterically block the binding of splicing factors to specific sites in pre-mRNAs or inhibit their function. Splicing activators enhance splicing by interacting with splicing factors, thus promoting their association with a regulated splice site.

■ In RNA editing the nucleotide sequence of a pre-mRNA is altered in the nucleus. In vertebrates, this process is fairly rare and entails deamination of a single base in the mRNA sequence, resulting in a change in the amino acid specified by the corresponding codon and production of a functionally different protein (see Figure 12-17).

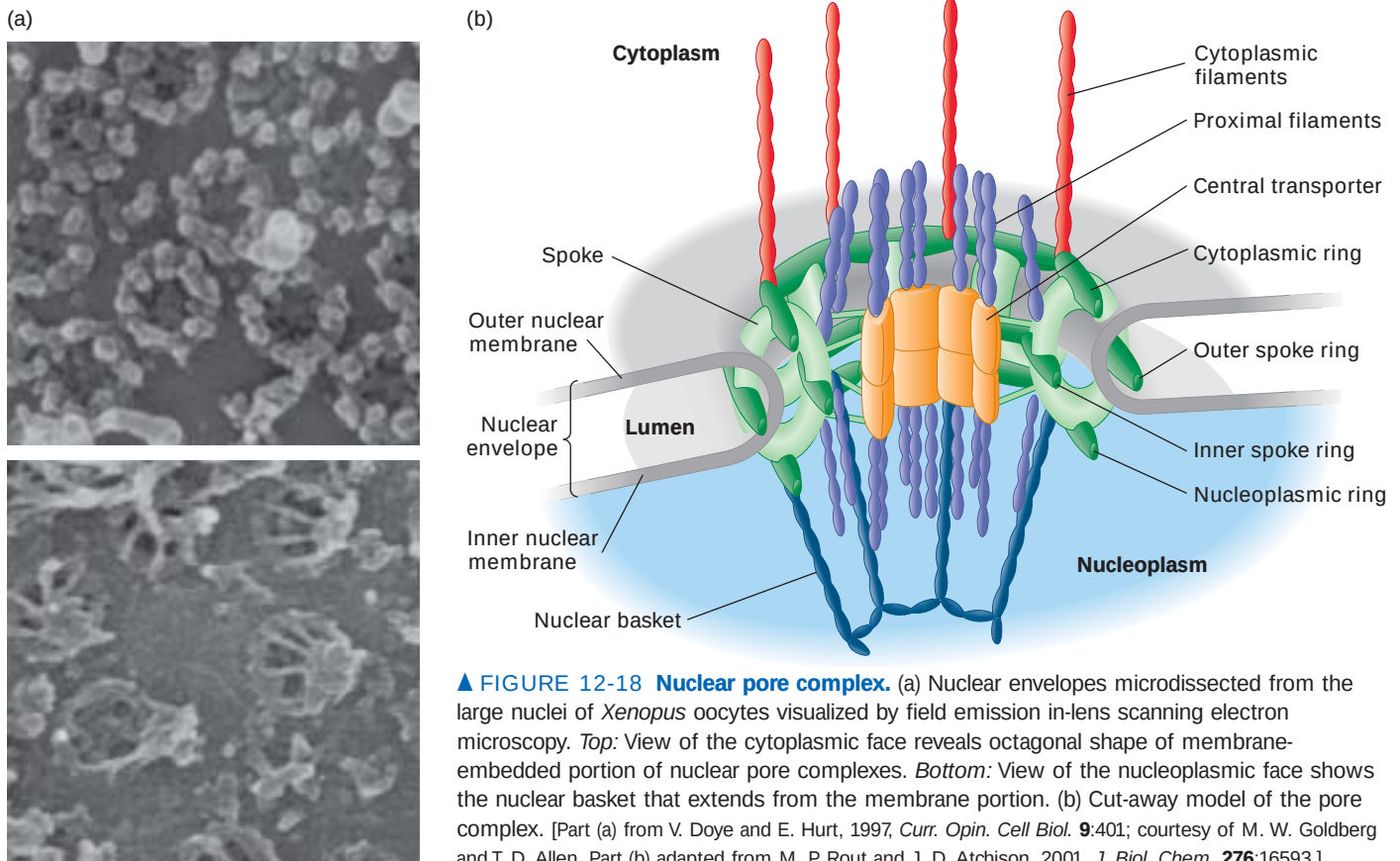
12.3 Macromolecular Transport Across the Nuclear Envelope

Once the processing of an mRNA is completed in the nucleus, it remains associated with specific hnRNP proteins in a *messenger ribonuclear protein complex*, or *mRNP*. Before it can be translated into the encoded protein, it must be exported out of the nucleus into the cytoplasm. The nucleus is separated from the cytoplasm by two membranes, which form the **nuclear envelope** (see Figure 5-19). Like the plasma membrane surrounding cells, each nuclear membrane consists of a water-impermeable phospholipid bilayer and vari-

ous associated proteins. Transport of macromolecules including mRNPs, tRNAs, and ribosomal subunits out of the nucleus and transport of all nuclear proteins translated in the cytoplasm into the nucleus occur through *nuclear pores* in a process that differs fundamentally from the transport of small molecules and ions across other cellular membranes (Chapter 7). Insight into the mechanisms of transport through nuclear pores first came from studies of the nuclear pores and of import and export of individual proteins, which we consider first, before returning to the export of mRNPs.

Large and Small Molecules Enter and Leave the Nucleus via Nuclear Pore Complexes

Numerous pores perforate the nuclear envelope in all eukaryotic cells. Each nuclear pore is formed from an elaborate structure termed the **nuclear pore complex (NPC)**, which is immense by molecular standards, ≈ 125 million daltons in vertebrates, or about 30 times larger than a ribosome. An NPC is made up of multiple copies of some 50 (in yeast) to 100 (in vertebrates) different proteins called **nucleoporins**. Electron micrographs of nuclear pore complexes reveal a roughly octagonal, membrane-embedded structure from which eight ≈ 100 -nm-long filaments extend into the nucleoplasm (Figure 12-18). The distal ends of these filaments are joined by the terminal ring, forming a structure called the



▲ **FIGURE 12-18 Nuclear pore complex.** (a) Nuclear envelopes microdissected from the large nuclei of *Xenopus* oocytes visualized by field emission in-lens scanning electron microscopy. *Top*: View of the cytoplasmic face reveals octagonal shape of membrane-embedded portion of nuclear pore complexes. *Bottom*: View of the nucleoplasmic face shows the nuclear basket that extends from the membrane portion. (b) Cut-away model of the pore complex. [Part (a) from V. Doye and E. Hurt, 1997, *Curr. Opin. Cell Biol.* **9**:401; courtesy of M. W. Goldberg and T. D. Allen. Part (b) adapted from M. P. Rout and J. D. Atchison, 2001, *J. Biol. Chem.* **276**:16593.]

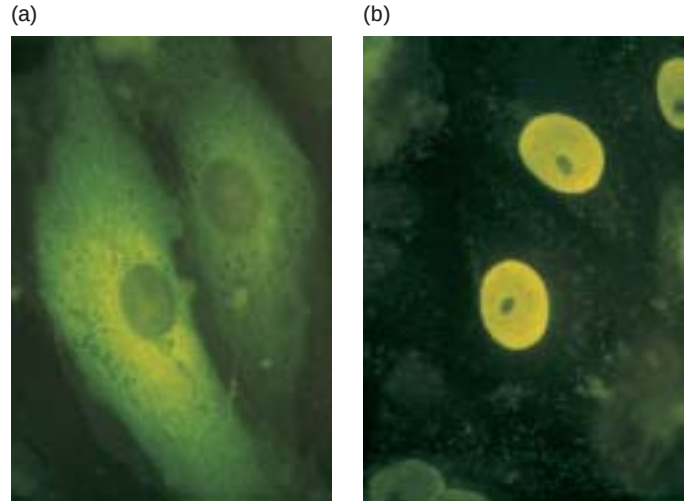
nuclear basket. The membrane-embedded portion also is attached directly to the **nuclear lamina**, a network of lamin intermediate filaments that extends over the inner surface of the nuclear envelope (see Figure 21-16). Cytoplasmic filaments extend from the cytoplasmic side of the NPC into the cytosol.

Ions, small metabolites, and globular proteins up to ≈ 60 kDa can diffuse through a water-filled channel in the nuclear pore complex; these channels behave as if they are ≈ 0.9 nm in diameter. However, large proteins and ribonucleoprotein complexes cannot diffuse in and out of the nucleus. Rather, these macromolecules are selectively transported in and out of the nucleus with the assistance of soluble transporter proteins that bind macromolecules and also interact with certain nucleoporins. The principles underlying macromolecular transport through nuclear pore complexes were first determined for the import of individual proteins into the nucleus, which we discuss first before turning to the question of how fully processed mRNAs are transported into the cytoplasm.

Importins Transport Proteins Containing Nuclear-Localization Signals into the Nucleus

All proteins found in the nucleus are synthesized in the cytoplasm and imported into the nucleus through nuclear pore complexes. Such proteins contain a *nuclear-localization signal (NLS)* that directs their selective transport into the nucleus. NLSs were first discovered through the analysis of mutants of simian virus 40 (SV40) that produced an abnormal form of the viral protein called large T-antigen. The wild-type form of this protein is localized to the nucleus in virus-infected cells, whereas some mutated forms of large T-antigen accumulate in the cytoplasm. The mutations responsible for this altered cellular localization all occur within a specific seven-residue sequence rich in basic amino acids near the C-terminus of the protein: Pro-Lys-Lys-Lys-Arg-Lys-Val. Experiments with engineered hybrid proteins in which this sequence was fused to a cytosolic protein demonstrated that it directs transport into the nucleus, and consequently functions as an NLS (Figure 12-19). NLS sequences subsequently were identified in numerous other proteins imported into the nucleus. Many of these are similar to the basic NLS in SV40 large T-antigen, whereas other NLSs are chemically quite different. For instance, an NLS in the hnRNP A1 protein is hydrophobic.

Early work on the mechanism of nuclear import focused on proteins containing a basic NLS, similar to the one in SV40 large T-antigen. A digitonin-permeabilized cell system provided an *in vitro* assay for analyzing soluble cytosolic components required for nuclear import (Figure 12-20). Using this assay system, four required proteins were purified: Ran, nuclear transport factor 2 (NTF2), importin α , and importin β . *Ran* is a monomeric **G protein** that exists in two conformations, one when complexed with GTP and an alternative one when the GTP is hydrolyzed to GDP (see Figure 3-29). The two **importins**

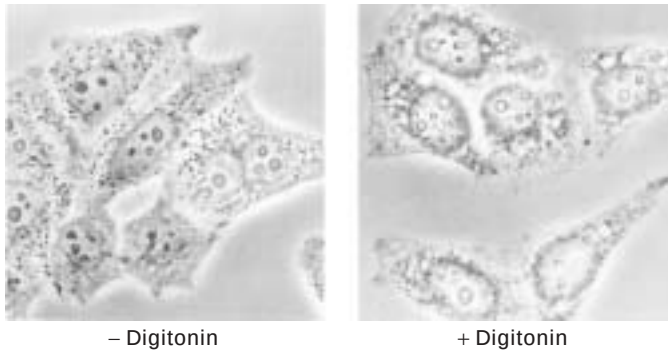


▲ **EXPERIMENTAL FIGURE 12-19 Fusion of a nuclear-localization signal (NLS) to a cytoplasmic protein causes the protein to enter the cell nucleus.** (a) Normal pyruvate kinase, visualized by immunofluorescence after treating cultured cells with a specific antibody (yellow), is localized to the cytoplasm. This very large cytosolic protein functions in carbohydrate metabolism. (b) When a chimeric pyruvate kinase protein containing the SV40 NLS at its N-terminus was expressed in cells, it was localized to the nucleus. The chimeric protein was expressed from a transfected engineered gene produced by fusing a viral gene fragment encoding the SV40 NLS to the pyruvate kinase gene. [From D. Kalderon et al., 1984, *Cell* **39**:499; courtesy of Dr. Alan Smith.]

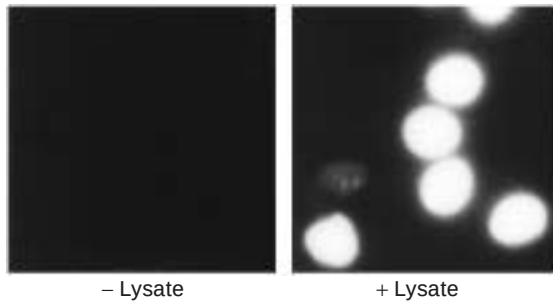
form a heterodimeric *nuclear-import receptor*: the α subunit binds to a basic NLS in a “cargo” protein to be transported into the nucleus, and the β subunit interacts with a class of nucleoporins called *FG-nucleoporins*. These nucleoporins, which line the channel of the nuclear pore complex and also are found in the nuclear basket and the cytoplasmic filaments, contain multiple repeats of short hydrophobic sequences rich in phenylalanine (F) and glycine (G) residues (FG-repeats). More recently, several importin β homologs have been found that function in the nuclear import of proteins with other classes of NLSs. These importins interact directly with their cognate NLSs without the need for an adapter protein like importin α .

A current model for the import of cytoplasmic cargo proteins mediated by a monomeric importin is shown in Figure 12-21. Free importin in the cytoplasm binds to its cognate NLS in a cargo protein, forming a *bimolecular cargo complex*. The cargo complex then translocates through the NPC channel as the importin binds transiently to successive individual FG-repeats in the FG-nucleoporins that line the channel. The FG-repeats are thought to act like “stepping stones” as the cargo complex diffuses from one FG-nucleoporin to another on its way through the channel by a process that does not require a direct input of energy from ATP hydrolysis or other mechanisms. The

(a) Effect of digitonin



(b) Nuclear import by permeabilized cells



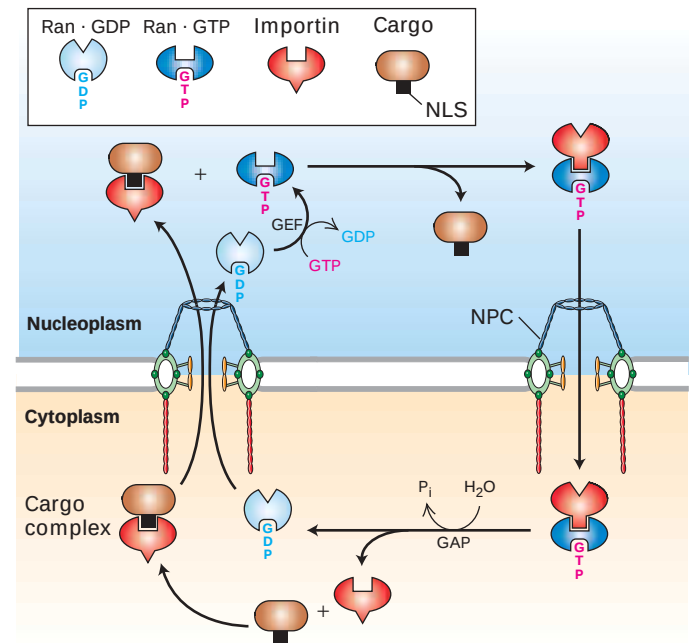
▲ **EXPERIMENTAL FIGURE 12-20** The failure of nuclear transport to occur in permeabilized cultured cells in the absence of lysate demonstrates the involvement of soluble cytosolic components in the process. (a) Phase-contrast micrographs of untreated and digitonin-permeabilized HeLa cells. Treatment of a monolayer of cultured cells with the mild, nonionic detergent digitonin permeabilizes the plasma membrane so that cytosolic constituents leak out, but leaves the nuclear envelope and NPCs intact. (b) Fluorescence micrographs of digitonin-permeabilized HeLa cells incubated with a fluorescent protein chemically coupled to a synthetic SV40 T-antigen NLS peptide in the presence and absence of cytosol (lysate). Accumulation of this transport substrate in the nucleus occurred only when cytosol was included in the incubation (*lower right*). [From S. Adam et al., 1990, *J. Cell. Biol.* **111**:807; courtesy of Dr. Larry Gerace.]

hydrophobic FG-repeats are thought to occur in regions of extended, otherwise hydrophilic polypeptide chains that fill the central transporter channel. The extended chains are proposed to form a meshwork of hydrophilic strands associated through hydrophobic interactions between the FG-repeats. This molecular meshwork is thought to restrict the diffusion of proteins larger than ≈ 60 kD. However, importins carrying cargo proteins are thought to penetrate the molecular meshwork by making successive interactions with the FG-repeats.

When the cargo complex reaches the nucleoplasm, the importin interacts with Ran·GTP, causing a conformational change in the importin that decreases its affinity for the NLS, releasing the cargo protein into the nucleoplasm.

The importin-Ran·GTP complex then diffuses back through the NPC, again, through transient interactions of the importin with FG repeats. Once the importin-Ran·GTP complex reaches the cytoplasmic side of the NPC, Ran interacts with a specific *GTPase-accelerating protein* (Ran-GAP) that is a component of the NPC cytoplasmic filaments. This stimulates Ran to hydrolyze its bound GTP to GDP, causing it to convert to a conformation that has low affinity for the importin, so that the free importin is released into the cytoplasm, where it can participate in another cycle of import.

Ran is returned to the nucleus by NTF2. The NTF2 dimer binds specifically to Ran·GDP and also interacts with the FG repeats of FG-nucleoporins. Consequently, the NTF2-Ran·GDP complex can diffuse through the pore via transient



▲ **FIGURE 12-21** Mechanism for nuclear import of “cargo” proteins. In the cytoplasm (*bottom*), a free importin binds to the NLS of a cargo protein, forming a bimolecular cargo complex. In the case of a basic NLS, the adapter protein importin α bridges the NLS and importin β , forming a trimolecular cargo complex (not shown). The cargo complex diffuses through the NPC by interacting with successive FG-nucleoporins. In the nucleoplasm, interaction of Ran·GTP with the importin causes a conformational change that decreases its affinity for the NLS, releasing the cargo. To support another cycle of import, the importin-Ran·GTP complex is transported back to the cytoplasm. A GTPase-accelerating protein (GAP) associated with the cytoplasmic filaments of the NPC stimulates Ran to hydrolyze the bound GTP. This generates a conformational change causing dissociation from the importin, which can then initiate another round of import. Ran·GDP is bound by NTF2 (not shown) and returned to the nucleoplasm, where a guanine nucleotide-exchange factor (GEF) causes release of GDP and rebinding of GTP.

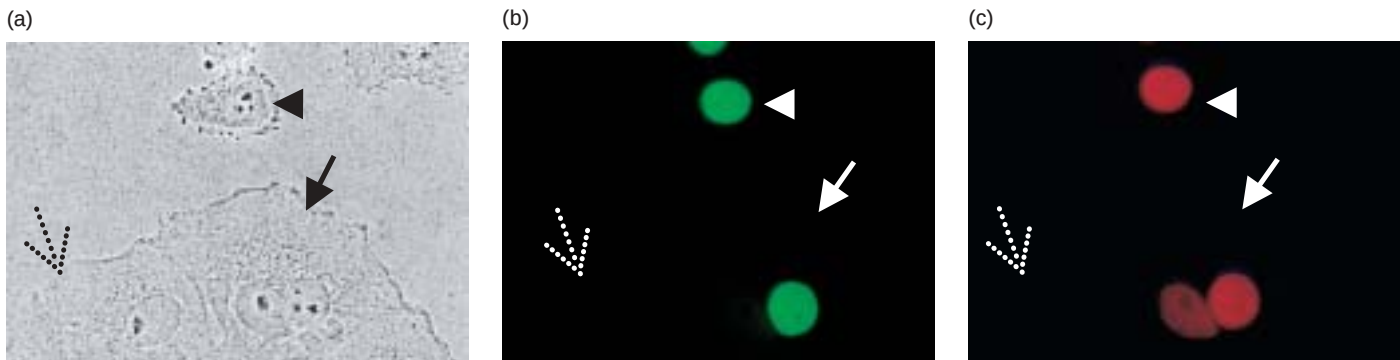
interactions between NTF2 and the FG repeats. When this complex reaches the nucleoplasm, it encounters a specific *guanine nucleotide–exchange factor* (*Ran-GEF*) that causes Ran to release its bound GDP and rebind GTP that is present at much higher concentration. The resulting conformational change in Ran decreases its affinity for NTF2 so that free Ran·GTP is released into the nucleoplasm and NTF2 is free to diffuse back through the pore.

The import complex travels through the pore by diffusion, a random process. Yet transport is unidirectional. The direction of transport is a consequence of the rapid dissociation of the import complex when it reaches the nucleoplasm. As a result, there is a concentration gradient of the importin-cargo complex across the NPC: high in the cytoplasm where the complex assembles and low in the nucleoplasm where it dissociates. This concentration gradient is responsible for the unidirectional nature of nuclear import. A similar concentration gradient is responsible for driving the importin in the nucleus back into the cytoplasm. The concentration of the importin-Ran·GTP complex is higher in the nucleoplasm, where it assembles, than on the cytoplasmic side of the NPC, where it dissociates. Ultimately, the direction of the transport processes is dependent on the asymmetric distribution of the Ran-GEF and the Ran-GAP. Ran-GEF in the nucleoplasm maintains Ran in the Ran·GTP state, where it promotes dissociation of the cargo complex. Ran-GAP on the cytoplasmic side of the NPC converts Ran·GTP to Ran·GDP, dissociating the importin-Ran·GTP complex and releasing free importin into the cytosol.

Exportins Transport Proteins Containing Nuclear-Export Signals out of the Nucleus

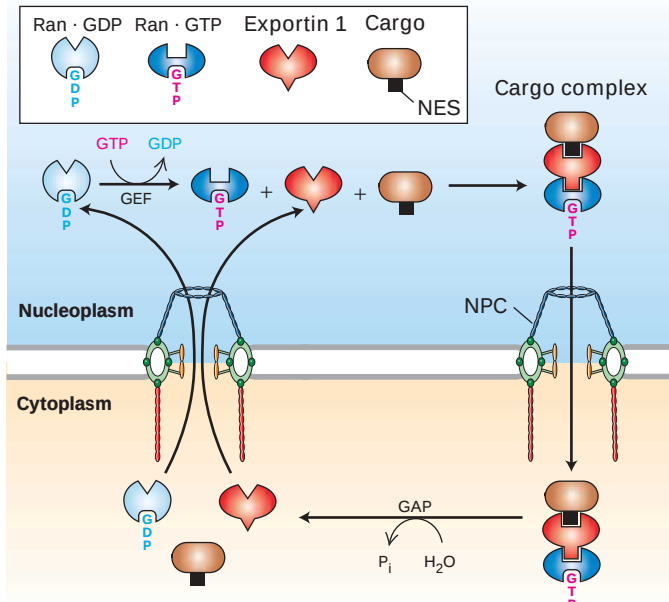
A very similar mechanism is used to export proteins, tRNAs, and ribosomal subunits from the nucleus to the cytoplasm. This mechanism initially was elucidated from studies of certain hnRNP proteins that “shuttle” between the nucleus and cytoplasm. For instance, the cell-fusion experiments described in Figure 12-22 first showed that some hnRNP proteins cycle in and out of the cytoplasm, whereas others remain localized in the nucleus. Such “shuttling” proteins contain a *nuclear-export signal* (*NES*) that stimulates their export from the nucleus to the cytoplasm through nuclear pores, in addition to an NLS that results in their reuptake into the nucleus. Experiments with engineered hybrid genes encoding a nucleus-restricted protein fused to various segments of a protein that shuttles in and out of the nucleus have identified at least three different classes of NESs: a leucine-rich sequence found in PKI (an inhibitor of protein kinase A) and in the Rev protein of human immunodeficiency virus (HIV), a 38-residue sequence in hnRNP A1, and a sequence in hnRNP K. To date, no functionally significant structural features (e.g., repeated amino acids) have been identified in the latter two classes.

The mechanism whereby shuttling proteins are exported from the nucleus is best understood for those containing a leucine-rich NES. According to the current model shown in Figure 12-23, a specific *nuclear-export receptor* in the nucleus, **exportin 1**, first forms a complex with Ran·GTP and then binds the NES in a cargo protein. Binding of exportin 1



▲ **EXPERIMENTAL FIGURE 12-22** The movement of human hnRNP A1 protein between nuclei in a heterokaryon shows that it can cycle in and out of the cytoplasm, but human hnRNP C protein, which showed no such movement, cannot. Cultured HeLa cells and *Xenopus* cells were fused by treatment with polyethylene glycol, producing heterokaryons containing nuclei from each cell type. The hybrid cells were treated with cycloheximide immediately after fusion to prevent protein synthesis. After 2 hours, the cells were fixed and stained with fluorescent-labeled antibodies specific for human hnRNP C and A1 proteins. These antibodies do not bind to the homologous *Xenopus* proteins. (a) A fixed preparation viewed by phase-contrast microscopy includes unfused HeLa cells (arrowhead) and *Xenopus* cells (dotted arrow), as well as fused heterokaryons

(solid arrow). In the heterokaryon in this micrograph, the round HeLa-cell nucleus is to the right of the oval-shaped *Xenopus* nucleus. (b, c) When the same preparation was viewed by fluorescence microscopy, the stained hnRNP C protein appeared green and the stained hnRNP A1 protein appeared red. Note that the unfused *Xenopus* cell on the left is unstained, confirming that the antibodies are specific for the human proteins. In the heterokaryon, hnRNP C protein appears only in HeLa-cell nuclei (b), whereas the A1 protein appears in both nuclei (c). Since protein synthesis was blocked after cell fusion, some of the human hnRNP A1 protein must have left the HeLa-cell nucleus, moved through the cytoplasm, and entered the *Xenopus* nucleus in the heterokaryon. [See S. Pinol-Roma and G. Dreyfuss, 1992, *Nature* **355**:730; courtesy of G. Dreyfuss.]



▲ **FIGURE 12-23 Mechanism for nuclear export of cargo proteins containing a leucine-rich nuclear-export signal (NES).** In the nucleoplasm, the protein exportin 1 binds cooperatively to the NES of the cargo protein to be transported and to Ran·GTP. After the resulting cargo complex diffuses through an NPC via transient interactions with FG-nucleoporins, the Ran GAP associated with the NPC cytoplasmic filaments stimulates conversion of Ran·GTP to Ran·GDP. The accompanying conformational change in Ran leads to dissociation of the complex. The NES-containing cargo protein is released into the cytosol, while exportin 1 and Ran·GDP are transported back into the nucleus through NPCs. Ran·GDP is transported through its interaction with NTF2. Ran·GEF in the nucleoplasm then stimulates conversion of Ran·GDP to Ran·GTP.

to Ran·GTP causes a conformational change in exportin 1 that increases its affinity for the NES so that a *trimolecular cargo complex* is formed. Like importins, exportin 1 interacts transiently with FG repeats in FG-nucleoporins and diffuses through the NPC. The cargo complex dissociates when it encounters the Ran-GAP in the NPC cytoplasmic filaments that stimulates Ran to hydrolyze the bound GTP, shifting it into a conformation that has low affinity for exportin 1. The released exportin 1 changes conformation to a structure that has low affinity for the NES, releasing the cargo into the cytosol. The direction of the export process is driven by this dissociation of the cargo from exportin 1 in the cytoplasm that causes a concentration gradient of the cargo complex across the NPC so that it is high in the nucleoplasm and low in the cytoplasm. Exportin 1 and the Ran·GDP are then transported back into the nucleus through an NPC. Ran·GDP is transported by NTF2, as discussed earlier in regard to nuclear import. It is then converted into Ran·GTP by the Ran-GEF in the nucleoplasm.

By comparing this model for nuclear export with that in Figure 12-21 for nuclear import, we can see one obvious dif-

ference: Ran·GTP is part of the cargo complex during export but not during import. Apart from this difference, the two transport processes are remarkably similar. In both processes, association of a transport signal receptor with Ran·GTP in the nucleoplasm causes a conformational change that affects its affinity for the transport signal. During import, the interaction causes release of the cargo, whereas, during export, the interaction promotes association with the cargo. In both export and import, stimulation of Ran·GTP hydrolysis in the cytoplasm by the Ran-GAP associated with NPC cytoplasmic filaments produces a conformational change in Ran that releases the transport signal receptor. During nuclear export, the cargo is also released. Importins and exportins both are thought to diffuse through the NPC channel by successive interactions with FG-repeats in FG-nucleoporins. Localization of the Ran-GAP and -GEF to the cytoplasm and nucleus, respectively, is the basis for the unidirectional transport of cargo proteins across the NPC.

In keeping with the similarity in function of importins and exportins, the two types of transport proteins are highly homologous in sequence and structure. The entire family is called the importin β family, or **karyopherins**. There are 14 karyopherins in yeast and more than 20 in mammalian cells. The NESs or NLSs to which they bind have been determined for only a fraction of them. Remarkably, some individual karyopherins function as both an importin and an exportin.

A similar shuttling mechanism has been shown to export other cargoes from the nucleus. For example, *exportin-t* functions to export tRNAs. Exportin-t binds fully processed tRNAs in a complex with Ran·GTP that diffuses through NPCs and dissociates when it interacts with Ran-GAP in the NPC cytoplasmic filaments, releasing the tRNA into the cytosol. A Ran-dependent process is also required for the nuclear export of ribosomal subunits through NPCs. Likewise, certain specific mRNAs that associate with particular hnRNP proteins (e.g., HIV Rev discussed later) can be exported by a Ran-dependent mechanism. However, most mRNAs are exported from the nucleus by another type of transporter that does not interact with Ran. This will be discussed after we examine gene regulation through control of transcription-factor import and export.

Control of Some Genes Is Achieved by Regulating Transport of Transcription Factors

Regulation of some genes is achieved by regulating the nuclear transport of specific transcription factors that control their transcription. For example, in Chapter 11 we learned that homodimeric **nuclear receptors** such as the glucocorticoid receptor are sequestered in the cytoplasm in the absence of ligand. Binding of ligand stimulates their translocation into the nucleus, where they can bind to hormone response elements in target genes and activate transcription (see Figures 11-43 and 11-44). The ligand-binding domain of these transcription factors contains an NLS, but in the absence of ligand, the NLS is masked by interactions with a cytoplasmic tethering complex. The conformational change that results

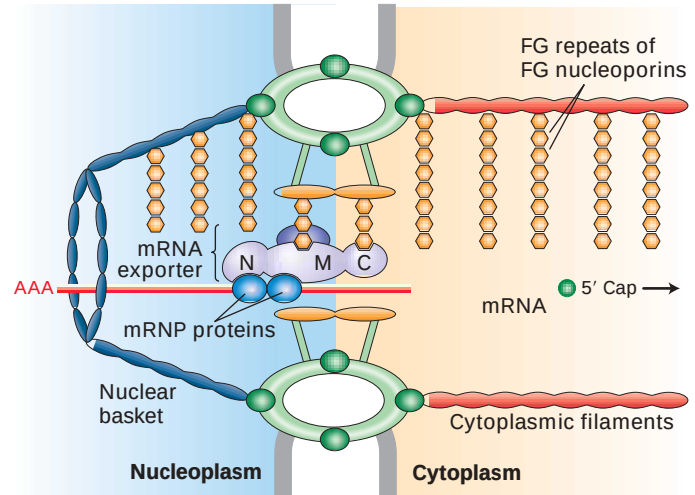
from binding hormone releases the domain from the tethering complex, exposing the NLS so that it is free to interact with its cognate importin and undergo nuclear import, as discussed above. Other examples of regulated nuclear import of transcription factors through controlled masking and exposure of their associated NLS are discussed in subsequent chapters on signal transduction.

In some cases, regulated nuclear transport of transcription factors is achieved by expressing them as fusions to domains that localize them to cytoplasmic membranes, such as the plasma membrane or ER. In response to specific signals, specific proteases are activated that cleave them, releasing the transcription-factor domains so that they can be transported into the nucleus by an importin. In other cases, transcription-factor nuclear export is regulated by post-translational modifications such as phosphorylation that increase their affinity for a specific exportin. Regulation of nuclear import or export is a powerful mechanism for controlling transcription-factor activity, since a transcription factor can bind to its cognate sites in the genes it regulates only when it is in the nucleus.

Most mRNPs Are Exported from the Nucleus with the Aid of an mRNA-Exporter

Recent studies of temperature-sensitive yeast mutants defective in nuclear transport identified a heterodimeric **mRNA-exporter** protein that appears to direct most mRNPs through nuclear pores. Yeast cells with mutations in the subunits of this protein accumulate polyadenylated RNA in the nucleus at the nonpermissive temperature, indicating that the protein is required for the transport of most mRNPs. Highly conserved homologs of the large and small subunits of the yeast mRNA-exporter are found in all eukaryotes, and several lines of evidence indicate that the mRNA-exporter is required for the export of most mRNPs in vertebrate cells as well as in yeast cells. Surprisingly, as mentioned earlier, experiments in both yeast and vertebrate cells indicate that Ran is not required for export of the vast majority of mRNPs.

The small subunit of the mRNA-exporter is homologous to NTF2 and interacts with a region of the larger subunit that also shares homology to NTF2. Together they form a domain that interacts with FG repeats in FG-nucleoporins similarly to NTF2, which is a dimer. A C-terminal domain of the large subunit also interacts with FG repeats. Although several other proteins are also required for mRNP transport, researchers consider this mRNA-exporter to be the primary transport protein because of the importance of FG-repeat binding for the mechanism of cargo transport by karyopherins. The large mRNA-exporter subunit also contains RNA-binding domains that appear to bind to RNA cooperatively with specific mRNP proteins. The finding that the mRNA-exporter is associated with mRNPs and binds directly to FG repeats led to the model of mRNP transport depicted in Figure 12-24. This model proposes that the mRNA-exporter translocates mRNPs through the nuclear pore channel similarly to the karyopherins, that is, by binding transiently to successive individual FG-repeats as it diffuses through the channel.



▲ **FIGURE 12-24 Proposed mechanism of mRNP transport from the nucleus via the mRNA-exporter.** The large subunit (light purple) of the mRNA-exporter contains three key domains: a middle domain (M) and a carboxyl domain (C), both of which bind to hydrophobic phenylalanine-glycine (FG) repeats in FG-nucleoporins, and an N-terminal region that has weak RNA-binding activity for most sequences. Binding to most mRNPs requires cooperative binding with other specific hnRNP proteins. The small subunit (dark purple) binds to the middle domain of the large subunit and contributes to the binding of FG repeats. For simplicity, FG repeats are not shown on the lower half of the nuclear pore complex. Note that the mRNA-exporter is not drawn to scale; it is far smaller than the nuclear pore complex. The mRNA-exporter is proposed to diffuse through the pore by making transient interactions with adjacent FG-nucleoporins as it progresses. [Adapted from R. Reed and E. Hurt, 2002, *Cell* **108**:523.]

Currently, it is not clear what provides directionality to this mechanism of mRNP transport through nuclear pores. By analogy with karyopherin transport, one hypothesis is that the mRNA-exporter-mRNP complex dissociates as it reaches the cytoplasm. This would result in a high concentration of the mRNA-exporter-mRNP complex at the nuclear basket, where it associates with FG-nucleoporins, and a low concentration of the complex in the cytoplasm, where it dissociates. As for karyopherin transport, such a concentration gradient could drive vectorial mRNP translocation. However, much remains to be learned about the hnRNP proteins associated with mRNPs during transport through the pore and the mechanism by which they dissociate from the mRNP on the cytoplasmic side of the pore.

Other Proteins That Assist in mRNP Export In addition to the mRNA-exporter and FG-nucleoporins, several other types of proteins are involved in the transport of mRNPs by this mechanism. As mentioned earlier, the mRNA-exporter is thought to bind to mRNAs cooperatively with specific mRNP proteins. For example, SR proteins associated with exons appear to stimulate the binding of the mRNA-exporter to processed mRNAs in mRNPs. Thus SR proteins not only

function to define exons during RNA splicing (see Figure 12-11) but also participate in the export process that translocates most mRNPs into the cytoplasm.

Other nuclear RNA-binding proteins that probably function in transport of mRNPs recently have been identified. These proteins, which initially were found associated with mRNAs spliced *in vitro*, bind to a region about 20 bases 5' of the exon sequences joined by splicing. Analysis of the proteins present in these *exon-junction complexes* revealed that several are homologous to yeast proteins that are altered in mutants defective in mRNP export. Moreover, one of these proteins has been shown to bind directly to the large subunit of the mRNA-exporter. These results suggest that exon-junction complexes deposited on mRNAs during RNA splicing stimulate mRNP transport through nuclear pores. Some of the proteins in exon-junction complexes also function as splicing factors. As we discuss in the next section, on mechanisms of cytoplasmic post-transcriptional control, the exon-junction complex also participates in a type of post-transcriptional control that prevents translation of improperly spliced mRNAs.

Another participant in mRNP transport to the cytoplasm is the nuclear cap-binding complex, mentioned earlier as protection against exonuclease attack on the 5' end of nascent transcripts and pre-mRNAs. Electron microscopy experiments discussed below have demonstrated that the 5' end of mRNAs lead the way through the nuclear pore complex. Recent experiments in yeast indicate that the 3' poly(A) tail plays an important role in mRNP transport, suggesting that a poly(A)-binding protein participates. Nucleoporins associated with the NPC cytoplasmic filaments in addition to FG-nucleoporins are required for mRNA export and may function to dissociate the mRNA-exporter and other mRNP proteins that accompany the mRNP through the pore.

Once the mRNP reaches the cytoplasm, most of the mRNP proteins that associated with the mRNA in the nucleus, the nuclear cap-binding complex, and the nuclear poly(A)-binding protein (PABPII) dissociate and are shuttled back to the nucleus. In the cytoplasm, the 5' cap of an exported mRNA is bound by the eIF4E translation initiation factor, the poly(A) tail is bound by multiple copies of the cytoplasmic poly(A)-binding protein (PABPI), and other RNA-binding proteins associate with the body of the mRNA, forming a cytoplasmic mRNP that has a lower ratio of protein to RNA than nuclear mRNPs.

Nuclear Export of Balbiani Ring mRNPs The salivary glands of larvae of the insect *Chironomus tentans* have provided a good model system for EM studies of the formation of hnRNPs and the export of mRNPs. In these larvae, genes in large chromosomal puffs called *Balbani rings* are abundantly transcribed into nascent pre-mRNAs that associate with hnRNP proteins and are processed into coiled mRNPs with an mRNA of ≈ 75 kb (Figure 12-25a, b). These giant mRNAs encode large glue proteins that adhere the developing larvae to a leaf. After processing of the pre-mRNA in Balbiani ring hnRNPs, the resulting mRNPs move through nuclear pores to

the cytoplasm. Electron micrographs of sections of these cells show mRNPs that appear to uncoil during their passage through nuclear pores and then bind to ribosomes as they enter the cytoplasm. The observation that mRNPs become associated with ribosomes during transport indicates that the 5' end leads the way through the nuclear pore complex, as mentioned earlier. Detailed electron microscopic studies of the transport of Balbiani ring mRNPs through nuclear pore complexes led to the model depicted in Figure 12-25c.

Pre-mRNAs in Spliceosomes Are Not Exported from the Nucleus

It is critical that only fully processed mature mRNAs be exported from the nucleus because translation of incompletely processed pre-mRNAs containing introns would produce defective proteins that might interfere with the functioning of the cell. By mechanisms that are not fully understood, pre-mRNAs associated with snRNPs in spliceosomes usually are prevented from being transported to the cytoplasm.

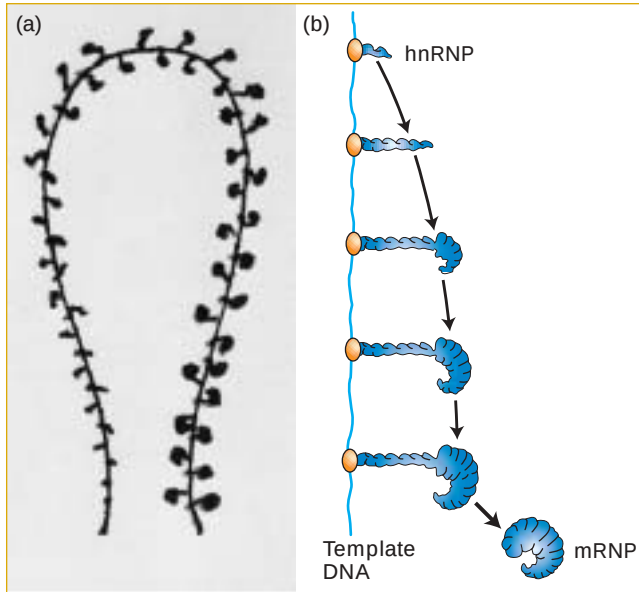
In one type of experiment demonstrating this restriction, a gene encoding a pre-mRNA with a single intron that normally is spliced out was mutated to introduce deviations from the consensus splice-site sequences. Mutation of *either* the 5' or the 3' invariant splice-site bases at the ends of the intron resulted in pre-mRNAs that were bound by snRNPs to form spliceosomes; however, RNA splicing was blocked, and the pre-mRNA was retained in the nucleus. In contrast, mutation of *both* the 5' and 3' splice sites in the same pre-mRNA resulted in export of the unspliced pre-mRNA, although less efficiently than for the spliced mRNA. When both splice sites were mutated, the pre-mRNAs were not efficiently bound by snRNPs, and, consequently, their export was not blocked.



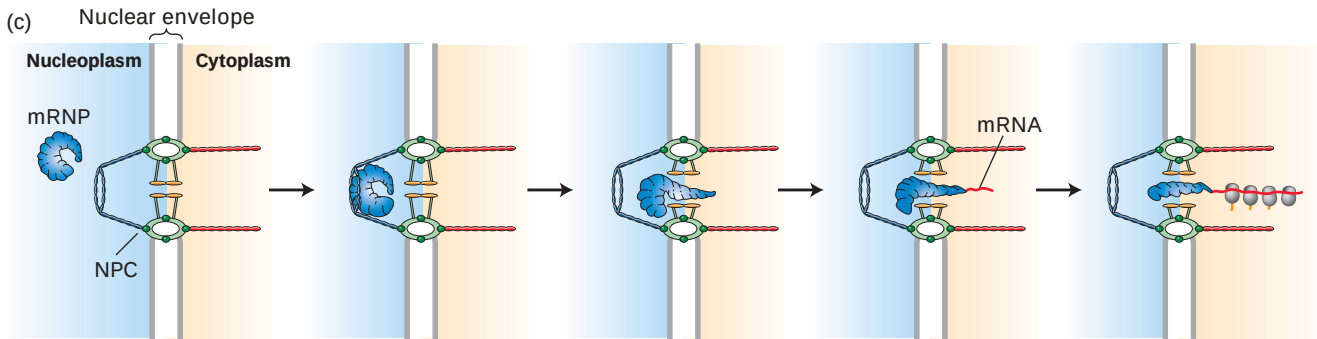
Many cases of thalassemia, an inherited disease that results in abnormally low levels of globin proteins, are due to mutations in globin-gene splice sites that decrease the efficiency of splicing but do not prevent association of the pre-mRNA with snRNPs. The resulting unspliced globin pre-mRNAs are retained in reticulocyte nuclei and are rapidly degraded. ■

HIV Rev Protein Regulates the Transport of Unspliced Viral mRNAs

As discussed earlier, transport of mRNPs containing mature, functional mRNAs from the nucleus to the cytoplasm entails a complex mechanism that is crucial to gene expression (see Figures 12-24 and 12-25). Regulation of this transport theoretically could provide another means of gene control, although it appears to be relatively rare. Indeed, the only examples of regulated mRNA export discovered to date occur during the cellular response to conditions (e.g., heat shock) that cause protein denaturation and during viral infection when virus-induced alterations in nuclear transport maximize viral replication. Here we describe the regulation



◀ **FIGURE 12-25 Formation of heterogeneous ribonucleoprotein particles (hnRNPs) and export of mRNPs from the nucleus.** (a) Model of a single chromatin transcription loop and assembly of Balbiani ring (BR) mRNP in *Chironomus tentans*. Nascent RNA transcripts produced from the template DNA rapidly associate with proteins, forming hnRNPs. The gradual increase in size of the hnRNPs reflects the increasing length of RNA transcripts at greater distances from the transcription start site. The model was reconstructed from electron micrographs of serial thin sections of salivary gland cells. (b) Schematic diagram of the biogenesis of hnRNPs. Following processing of the pre-mRNA, the resulting ribonucleoprotein particle is referred to as an mRNP. (c) Model for the transport of BR mRNPs through the nuclear pore complex (NPC) based on electron microscopic studies. Note that the curved mRNPs appear to uncoil as they pass through nuclear pores. As the mRNA enters the cytoplasm, it rapidly associates with ribosomes, indicating that the 5' end passes through the NPC first. [Part (a) from C. Errison et al., 1989, *Cell* **56**:631; courtesy of B. Daneholt. Parts (b) and (c) adapted from B. Daneholt, 1997, *Cell* **88**:585. See also B. Daneholt, 2001, *Proc. Nat'l. Acad. Sci. USA* **98**:7012.]



of mRNP export mediated by a protein encoded by human immunodeficiency virus (HIV).

A retrovirus, HIV integrates a DNA copy of its RNA genome into the host-cell DNA (see Figure 4-43). The integrated viral DNA, or provirus, contains a single transcription unit, which is transcribed into a single primary transcript by cellular RNA polymerase II. The HIV transcript can be spliced in alternative ways to yield three classes of mRNAs: a 9-kb unspliced mRNA; $\approx$ 4-kb mRNAs formed by removal of one intron; and $\approx$ 2-kb mRNAs formed by removal of two or more introns (Figure 12-26). After their synthesis in the host-cell nucleus, all three classes of HIV mRNAs are transported to the cytoplasm and translated into viral proteins; some of the 9-kb unspliced RNA is used as the viral genome in progeny virions that bud from the cell surface.

Since the 9-kb and 4-kb HIV mRNAs contain splice sites, they can be viewed as incompletely spliced mRNAs. However, as discussed earlier, association of such incompletely spliced mRNAs with snRNPs in spliceosomes normally blocks their export from the nucleus. Thus HIV, as well as other retroviruses, must have some mechanism for overcoming this block, permitting export of the longer viral mRNAs. Some retroviruses have evolved a sequence called the *constitutive transport element (CTE)* that binds to the mRNA-exporter with

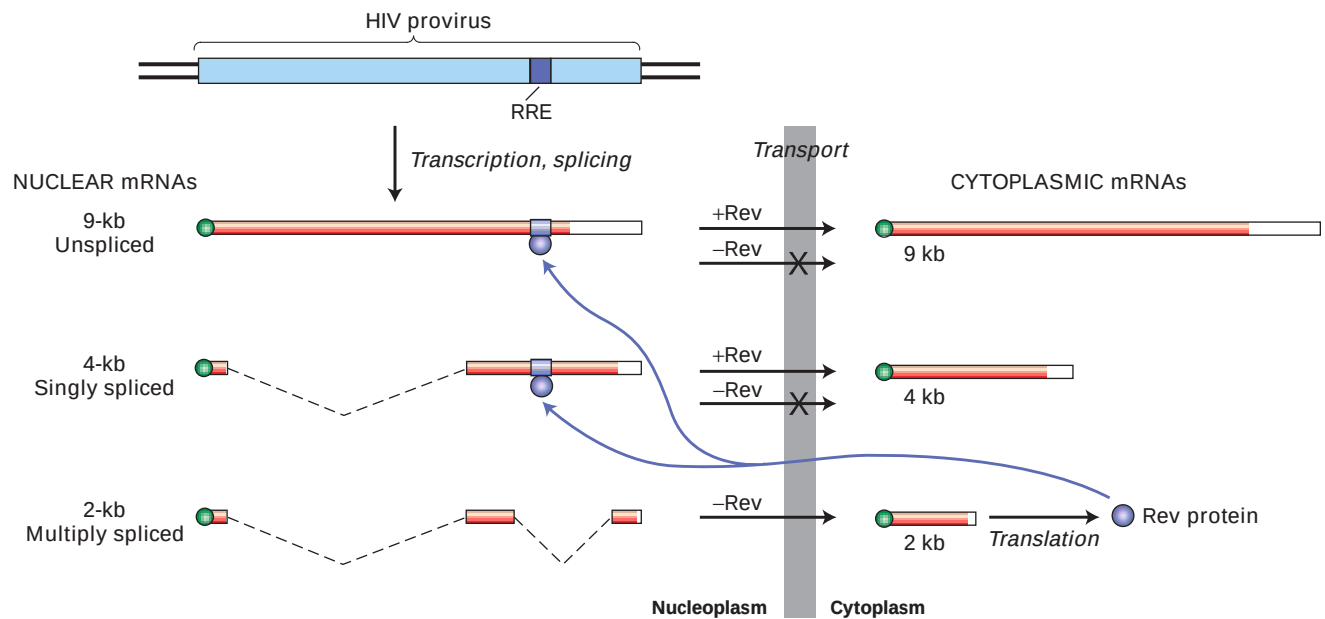
high affinity, thereby permitting export of unspliced retroviral RNA into the cytoplasm. HIV solved the problem differently.

Studies with HIV mutants showed that transport of unspliced 9-kb and singly spliced 4-kb viral mRNAs from the nucleus to the cytoplasm requires the virus-encoded *Rev* protein. Subsequent biochemical experiments demonstrated that Rev binds to a specific Rev-response element (RRE) present in HIV RNA. In cells infected with HIV mutants lacking the RRE, unspliced and singly spliced viral mRNAs remain in the nucleus, demonstrating that the RRE is required for Rev-mediated stimulation of transport. Rev contains a leucine-rich NES that interacts with exportin 1 complexed with Ran·GTP. As a result, Rev exports unspliced and singly spliced HIV mRNAs through interactions with exportin 1 and the nuclear pore complex.

KEY CONCEPTS OF SECTION 12.3

Macromolecular Transport Across the Nuclear Envelope

■ The nuclear envelope contains numerous nuclear pore complexes (NPCs), large, complicated structures composed of multiple copies of $\approx$ 50–100 proteins called *nucleoporins* (see Figure 12-18). FG-nucleoporins, which contain multiple



▲ **FIGURE 12-26 Role of Rev protein in transport of HIV mRNAs from the nucleus to the cytoplasm.** The HIV genome, which contains several coding regions, is transcribed into a single 9-kb primary transcript. Several ≈4-kb mRNAs result from alternative splicing out of any one of several introns (dashed lines), and several ≈2-kb mRNAs from splicing out of two or

more alternative introns. After transport to the cytoplasm, the various RNA species are translated into different viral proteins. Rev protein, encoded by a 2-kb mRNA, interacts with the Rev-response element (RRE) in the unspliced and singly spliced mRNAs, stimulating their transport to the cytoplasm. [Adapted from B. R. Cullen and M. H. Malim, 1991, *Trends Biochem. Sci.* **16**:346.]

repeats of a short hydrophobic sequence (FG-repeats), line the central transporter channel and play a role in transport of all macromolecules through nuclear pores.

- Transport of macromolecules larger than ≈60 kDa through nuclear pores requires the assistance of proteins that interact with both the transported molecule and with FG-repeats of FG-nucleoporins.

- Proteins imported to or exported from the nucleus contain a specific amino acid sequence that functions as a nuclear-localization signal (NLS) or a nuclear-export signal (NES). Nucleus-restricted proteins contain an NLS but not an NES, whereas proteins that shuttle between the nucleus and cytoplasm contain both signals.

- Several different types of NES and NLS have been identified. Each type of nuclear-transport signal is thought to interact with a specific receptor protein (exportin or importin) belonging to a family of homologous proteins termed *karyopharins*.

- A “cargo” protein bearing an NES or NLS translocates through nuclear pores bound to its cognate receptor protein (karyopharin), which also interacts with FG-nucleoporins. Importins and exportins are thought to diffuse through the channel by binding transiently to different FG-repeats that act like “stepping stones” through the pore. Both transport processes also require participation of Ran, a monomeric G protein that exists in different conformations when bound to GTP or GDP.

- After a cargo complex reaches its destination (the cytoplasm during export and the nucleus during import), it dissociates, freeing the cargo protein and other components. The latter then are transported through nuclear pores in the reverse direction to participate in transporting additional molecules of cargo protein (see Figures 12-21 and 12-23).

- The unidirectional nature of protein export and import through nuclear pores results from localization of the Ran guanine nucleotide-exchange factor (GEF) in the nucleus and of Ran GTPase-accelerating protein (GAP) in the cytoplasm. The interaction of import cargo complexes with the Ran-GEF in the nucleoplasm causes dissociation of the complex, releasing the cargo into the nucleoplasm (see Figure 12-21). Export cargo complexes dissociate in the cytoplasm when they interact with Ran GAP localized to the NPC cytoplasmic filaments (see Figure 12-23).

- Most mRNPs are exported from the nucleus by a heterodimeric mRNA-exporter that interacts with FG-repeats (see Figure 12-24). The direction of transport (nucleus → cytoplasm) may result from dissociation of the exporter-mRNP complex in the cytoplasm by an as yet uncharacterized mechanism that does not depend on Ran.

- The mRNA-exporter binds to most mRNAs cooperatively with SR proteins bound to exons, to exon-junction complexes that associate with mRNAs following RNA

splicing, and to additional mRNP proteins that remain to be characterized.

- Pre-mRNAs bound by a spliceosome normally are not exported from the nucleus, assuring that only fully processed, functional mRNAs reach the cytoplasm for translation.

12.4 Cytoplasmic Mechanisms of Post-transcriptional Control

Before proceeding, let's quickly review the steps in gene expression at which control is exerted. We saw in the previous chapter that regulation of transcription initiation is the principal mechanism for controlling expression of genes. In preceding sections of this chapter we learned that transcription of some genes is regulated by controlling the transport into and out of the nucleus of transcription factors that regulate them. We also learned that expression of protein isoforms is controlled by regulating alternative RNA splicing. Although transport of fully processed mRNAs to the cytoplasm rarely is regulated, the transport of unspliced retroviral RNAs is specifically controlled.

In this section we consider other mechanisms of post-transcriptional control that contribute to regulating the expression of some genes. Most of these mechanisms operate in the cytoplasm, controlling the stability or localization of mRNA or its translation into protein. We begin by discussing two recently discovered and related mechanisms of gene control.

Micro RNAs Repress Translation of Specific mRNAs

Micro RNAs (miRNAs) were first discovered during analysis of mutations in the *lin-4* and *let-7* genes of the nematode *C. elegans*. Cloning and analysis of wild-type *lin-4* and *let-7* revealed that they encode no protein products, but rather RNAs only 21 and 22 nucleotides long, respectively, that hybridize to the 3' untranslated regions of specific target mRNAs. For example, the *lin-4* miRNA, which is expressed early in embryogenesis, hybridizes to the 3' untranslated regions of both the *lin-14* and *lin-28* mRNAs, thereby repressing translation of these mRNAs by an as yet unknown mechanism. Expression of *lin-4* miRNA ceases later in development, allowing translation of newly synthesized *lin-14* and *lin-28* mRNAs at that time. Expression of *let-7* miRNA occurs at comparable times during embryogenesis of all bilaterally symmetric animals. The role of *lin-4* and *let-7* miRNAs in coordinating the timing of early developmental events in *C. elegans* is discussed in Chapter 22. Here we focus on what is currently understood about how miRNAs repress translation.

This form of translational regulation is probably not limited to the *lin-4* and *let-7* miRNAs in *C. elegans*; about 100 different miRNAs have been found in *C. elegans*, and at least

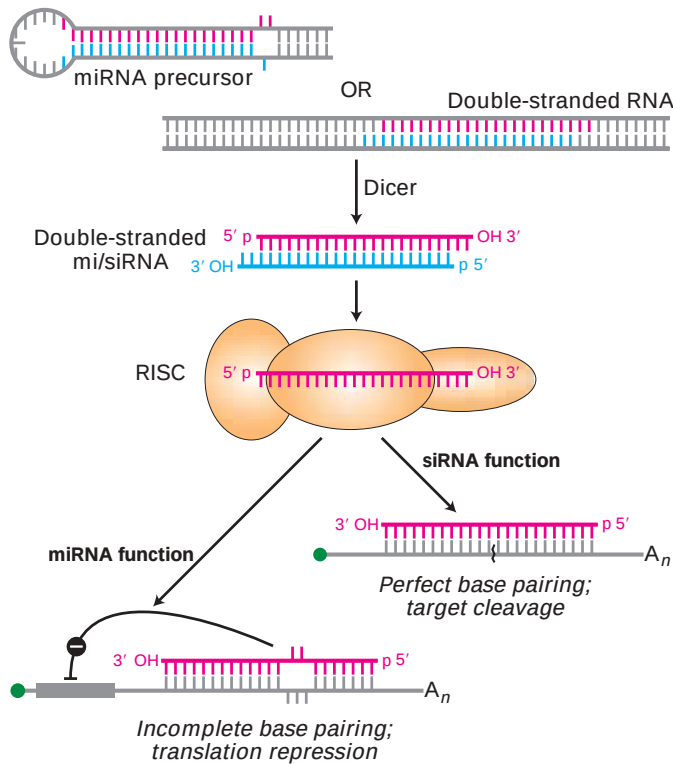
as many in humans. All miRNAs appear to be formed by processing of ≈ 70 -nucleotide precursor RNAs that form hairpin structures with a few base-pair mismatches in the stem of the hairpin. A ribonuclease called *Dicer*, which cleaves double-stranded RNA, is required for production of miRNAs from these precursors. The base pairing between a miRNA and the 3' untranslated region of its target mRNAs is not precisely complementary, so that some base-pair mismatches occur in the hybridized region. This mismatching distinguishes miRNA-mediated translational repression from the related phenomenon of RNA interference, which we describe next.

RNA Interference Induces Degradation of mRNAs with Sequences Complementary to Double-Stranded RNAs

RNA interference (RNAi) was discovered unexpectedly during attempts to experimentally manipulate the expression of specific genes. Researchers tried to inhibit the expression of a gene in *C. elegans* by microinjecting a single-stranded, complementary RNA that would hybridize to the encoded mRNA and prevent its translation, a method called *antisense inhibition*. But in control experiments, perfectly base-paired double-stranded RNA a few hundred base pairs long was much more effective at inhibiting expression of the gene than the antisense strand alone. Similar inhibition of gene expression by an introduced double-stranded RNA soon was observed in plants. In each case, the double-stranded RNA induced degradation of all cellular RNAs containing a sequence that was exactly the same as one strand of the double-stranded RNA. Because of the specificity of RNA interference in targeting mRNAs for destruction, it has become a powerful experimental tool for studying gene function (see Figure 9-43).

Subsequent biochemical studies with extracts of *Drosophila* embryos showed that a long double-stranded RNA that mediates interference is initially processed into a double-stranded intermediate referred to as *short interfering RNA (siRNA)*. The strands in siRNA contain 21–23 nucleotides hybridized to each other so that the two bases at the 3' end of each strand are single-stranded. The finding that Dicer ribonuclease is required for formation of siRNAs suggested that RNA interference and miRNA-mediated translational repression are related processes.

Recent studies indicate that double-stranded siRNAs and miRNAs are further processed into a multiprotein complex containing only one of the RNA strands (Figure 12-27). This **RNA-induced silencing complex (RISC)** then cleaves target RNAs that are *precisely* complementary to their corresponding single-stranded siRNAs. These complexes also appear to function in the inhibition of translation by miRNAs. The human *let-7* miRNA, for instance, is found in an RNA-induced silencing complex that can cleave a synthetic target RNA that is precisely complementary to *let-7* miRNA. However, the same complex does not cleave an RNA whose



▲ **FIGURE 12-27 Model for miRNA translational repression and RNA interference mediated by the RNA-induced interfering complex (RISC).** Processing of both miRNA precursors into miRNAs and long double-stranded RNAs into short interfering RNAs (siRNAs) requires the Dicer ribonuclease. In both cases, cleavage by Dicer yields a double-stranded RNA intermediate containing 21–23 nucleotides per strand with two-nucleotide 3' single-stranded tails. One strand of this intermediate assembles with multiple proteins to form an RNA-induced silencing complex (RISC). In siRNA function, the target mRNA hybridizes perfectly with the RISC RNA, leading to cleavage of the target. In miRNA function, the RISC RNA forms a hybrid with the target mRNA that contains some base-pair mismatches; in this case translation of the target mRNA is blocked. This is the situation for the *lin-4* and *let-7* miRNAs in *C. elegans*. [Adapted from G. Hutvagner and P. D. Zamore, 2002, *Science* **297**:2056; see also V. Ambrose, 2001, *Cell* **107**:823.]

sequence differs at a few bases from the *let-7* sequence. This finding has led to the hypothesis that RISC complexes have two distinct functions: (1) an siRNA function (i.e., RNA interference) leading to cleavage of precisely complementary mRNAs and (2) an miRNA function leading to translational repression of mRNAs with a few base-pair mismatches (see Figure 12-27).



Mutations in the fragile-X gene (*FMR1*), which encodes a protein containing a KH RNA-binding motif, are associated with the most common form of heritable mental retardation. Recent results indicate that the *FMR1* protein is a subunit of human RISC complexes.

Consequently, the mental retardation associated with *FMR1* mutations may result from a defect in translational repression of specific target mRNAs during development of the central nervous system. ■

RNA interference is believed to be an ancient cellular defense against certain viruses and mobile genetic elements in both plants and animals. Plants with mutations in the genes encoding Dicer and RISC proteins exhibit increased sensitivity to infection by RNA viruses and increased movement of **transposons** within their genomes. The double-stranded RNA intermediates generated during replication of RNA viruses are thought to be recognized by the Dicer ribonuclease, inducing an RNAi response that ultimately degrades viral mRNAs. During transposition, transposons are inserted into cellular genes in a random orientation, and their transcription from different promoters produces complementary RNAs that can hybridize with each other, initiating the RNAi system that then interferes with the expression of transposon proteins required for additional transpositions.

In plants and *C. elegans* the RNAi response can be induced in all cells of the organism by introduction of double-stranded RNA into just a few cells. Such organism-wide induction requires a protein produced in plants and in *C. elegans* that is homologous to the RNA replicases of RNA viruses. This finding suggests that double-stranded siRNAs are replicated and then transferred to other cells in these organisms. In plants, transfer of siRNAs might occur through **plasmodesmata**, the cytoplasmic connections between plant cells that traverse the cell walls between them (see Figure 6-34). Organism-wide induction of RNA interference does not occur in *Drosophila* or mammals, presumably because their genomes do not encode RNA replicase homologs. In plants, double-stranded RNA also induces the DNA methylation of genes with the same sequence by an unknown mechanism. Such gene methylation in response to RNAi inhibits transcription of the gene, probably through the binding of histone deacetylases. This type of RNAi-induced gene methylation does not occur in animals.

Cytoplasmic Polyadenylation Promotes Translation of Some mRNAs

In addition to repression of translation by micro RNAs, *protein-mediated* translational control helps regulate expression of some genes. Regulatory sequences, or elements, in mRNAs that interact with specific proteins to control translation generally are present in the untranslated region (UTR) at the 3' or 5' end of an mRNA. Here we discuss a type of protein-mediated translational control involving 3' regulatory elements. A different mechanism involving RNA-binding proteins that interact with 5' regulatory elements is discussed later.

Sequence-specific translation-control proteins may bind cooperatively to neighboring sites in 3' UTRs and function in a combinatorial manner, similar to the cooperative binding of transcription factors to regulatory sites in an enhancer

region of the synapse, leading to the local synthesis of new proteins that increase the size and alter the neurophysiological characteristics of the synapse. The finding that CPEB is present in neuronal dendrites has led to the proposal that cytoplasmic polyadenylation stimulates translation of specific mRNAs in dendrites, much as it does in oocytes. In this case, presumably, synaptic activity (rather than a hormone) is the signal that induces phosphorylation of CPEB and subsequent activation of translation.

mRNAs Are Degraded by Several Mechanisms in the Cytoplasm

The concentration of an mRNA is a function of both its rate of synthesis and its rate of degradation. For this reason, if two genes are transcribed at the same rate, the steady-state concentration of the corresponding mRNA that is more stable will be higher than the concentration of the other. The stability of an mRNA also determines how rapidly synthesis of the encoded protein can be shut down. For a stable mRNA, synthesis of the encoded protein persists long after transcription of the gene is repressed. Most bacterial mRNAs are unstable, decaying exponentially with a typical half-life of a few minutes. For this reason, a bacterial cell can rapidly adjust the synthesis of proteins to accommodate changes in the cellular environment. Most cells in multicellular organisms, on the other hand, exist in a fairly constant environment and carry out a specific set of functions over periods of days to months or even the lifetime of the organism (nerve cells, for example). Accordingly, most mRNAs of higher eukaryotes have half-lives of many hours.

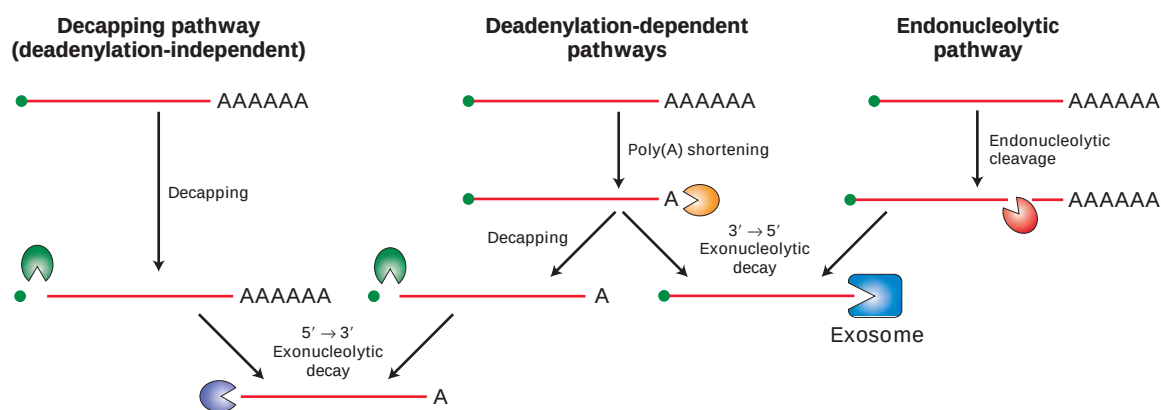
However, some proteins in eukaryotic cells are required only for short periods of time and must be expressed in bursts. For example, certain signaling molecules called **cy-**

tokines, which are involved in the immune response of mammals, are synthesized and secreted in short bursts. Similarly, many of the transcription factors that regulate the onset of the S phase of the cell cycle, such as c-Fos and c-Jun, are synthesized for brief periods only (Chapter 21). Expression of such proteins occurs in short bursts because transcription of their genes can be rapidly turned on and off and their mRNAs have unusually short half-lives, on the order of 30 minutes or less.

Cytoplasmic mRNAs are degraded by one of the pathways shown in Figure 12-29. For most mRNAs, the length of the poly(A) tail gradually decreases with time through the action of a deadenylating nuclease. When it is shortened sufficiently, PABPI molecules can no longer bind and stabilize interaction of the 5' cap and initiation factors (see Figure 4-31). The exposed cap then is removed by a decapping enzyme, and the unprotected mRNA is degraded by a 5' → 3' exonuclease. Removal of the poly(A) tail also makes mRNAs susceptible to degradation by cytoplasmic exosomes containing 3' → 5' exonucleases. The 5' → 3' exonucleases predominate in yeast, and the 3' → 5' exosome apparently predominates in mammalian cells.

For mRNAs degraded in these deadenylation-dependent pathways, the rate at which they are deadenylated controls the rate at which they are degraded. The rate of deadenylation varies inversely with the frequency of translation initiation for an mRNA: the higher the frequency of initiation, the slower the rate of deadenylation. This relation probably is due to the reciprocal interactions between initiation factors and PABPI that stabilize the binding of PABPI to the poly(A) tail, thereby protecting it from the deadenylation exonuclease.

Many short-lived mRNAs in mammalian cells contain multiple, sometimes overlapping, copies of the sequence AUUUA in their 3' untranslated region. Specific RNA-binding



▲ **FIGURE 12-29 Pathways for degradation of eukaryotic mRNAs.** In the deadenylation-dependent (*middle*) pathways, the poly(A) tail is progressively shortened by a deadenylase (orange) until it reaches a length of 20 or fewer A residues at which the interaction with PABPI is destabilized, leading to weakened interactions between the 5' cap and translation initiation factors. The deadenylated mRNA then may either (1) be decapped and

degraded by a 5' → 3' exonuclease or (2) be degraded by a 3' → 5' exonuclease in cytoplasmic exosomes. Some mRNAs (*right*) are cleaved internally by an endonuclease, and the fragments degraded by an exosome. Other mRNAs (*left*) are decapped before they are deadenylated, and then degraded by a 5' → 3' exonuclease. [Adapted from M. Tucker and R. Parker, 2000, *Ann. Rev. Biochem.* 69:571]

proteins have been found to bind to these 3' AU-rich sequences. Recent experiments suggest that the bound proteins interact with a deadenylating enzyme and with the exosome, thereby promoting the rapid deadenylation and subsequent 3' → 5' degradation of these mRNAs. In this mechanism, the rate of mRNA degradation is uncoupled from the frequency of translation. Thus mRNAs containing the AUUUA sequence can be translated at high frequency, yet also degraded rapidly, allowing the encoded proteins to be expressed in short bursts.

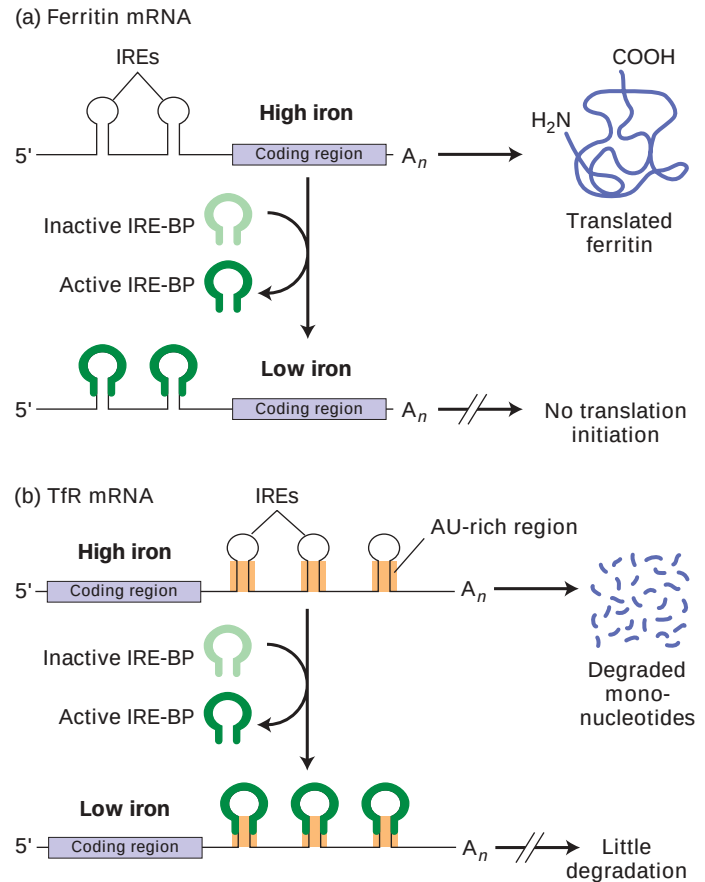
As shown in Figure 12-29, some mRNAs are degraded in pathways that do not involve significant deadenylation. In one of these, mRNAs are decapped before the poly(A) tail is shortened extensively. It appears that certain mRNA sequences make the cap sensitive to the decapping enzyme, but the precise mechanism is unclear. In the other alternative pathway, mRNAs first are cleaved internally by endonucleases. The RNA-induced silencing complex (RISC) discussed earlier is an example of such an endonuclease (see Figure 12-27). The fragments generated by internal cleavage then are degraded by exonucleases.

An Iron-Sensitive RNA-Binding Protein Regulates mRNA Translation and Degradation

Control of intracellular iron concentrations by the *iron-response element-binding protein* (IRE-BP) is an elegant example of a single protein that regulates the translation of one mRNA and the degradation of another. When intracellular iron stores are low, this dual control system operates to increase the level of free iron ions available for iron-requiring enzymes; when iron is in excess, the system operates to prevent accumulation of toxic levels of free ions. It is one of the simplest and best-understood examples of protein-mediated translational control.

One component in this system is the regulation of production of *ferritin*, an intracellular iron-binding protein. The 5' untranslated region of ferritin mRNA contains *iron-response elements* (IREs) that have a stem-loop structure. IRE-BP recognizes five specific bases in the IRE loop and the duplex nature of the stem. At low iron concentrations, IRE-BP is in an active conformation that binds to the IREs (Figure 12-30a). The bound IRE-BP blocks the 40S ribosomal subunit from scanning for the AUG start codon (see Figure 4-25), thereby inhibiting translation initiation. The resulting decrease in ferritin means less iron is complexed with the ferritin and is therefore available to iron-requiring enzymes. At high iron concentrations, IRE-BP is in an inactive conformation that does not bind to the 5' IREs, so translation initiation can proceed. The newly synthesized ferritin then binds free iron ions, preventing their accumulation to harmful levels.

The other part of this regulatory system controls the import of iron into cells. In vertebrates, ingested iron is carried through the circulation bound to a protein called *transferrin*. After binding to the transferrin receptor (TfR) in the plasma membrane, the transferrin-iron complex is brought into cells



▲ **FIGURE 12-30 Iron-dependent regulation of mRNA translation and degradation.** The iron response element-binding protein (IRE-BP) controls translation of ferritin mRNA (a) and degradation of transferrin-receptor (TfR) mRNA (b). At low intracellular iron concentrations IRE-BP binds to iron-response elements (IREs) in the 5' or 3' untranslated region of these mRNAs. At high iron concentrations, IRE-BP undergoes a conformational change and cannot bind either mRNA. The dual control by IRE-BP precisely regulates the level of free iron ions within cells. See the text for discussion.

by receptor-mediated endocytosis (Chapter 17). The 3'-untranslated region of TfR mRNA contains IREs whose stems have AU-rich destabilizing sequences (Figure 12-30b). At high iron concentrations, when the IRE-BP is in the inactive, nonbinding conformation, these AU-rich sequences are thought to promote degradation of TfR mRNA by the same mechanism that leads to rapid degradation of other short-lived mRNAs, as described previously. The resulting decrease in production of the transferrin receptor quickly reduces iron import, thus protecting the cell. At low iron concentrations, however, IRE-BP can bind to the 3' IREs in TfR mRNA. The bound IRE-BP is thought to block recognition of the destabilizing AU-rich sequences by the proteins that would otherwise rapidly degrade the mRNAs. As a result, production of the transferrin receptor increases and more iron is brought into the cell.

Other regulated RNA-binding proteins may also function to control mRNA translation or degradation, much like the dual-acting IRE-BP. For example, a heme-sensitive RNA-binding protein controls translation of the mRNA encoding aminolevulinate (ALA) synthase, a key enzyme in the synthesis of heme. And *in vitro* studies have shown that the mRNA encoding the milk protein casein is stabilized by the hormone prolactin and rapidly degraded in its absence.

Nonsense-Mediated Decay and Other mRNA Surveillance Mechanisms Prevent Translation of Improperly Processed mRNAs

Translation of an improperly processed mRNA could lead to production of an abnormal protein that interferes with functioning of the normal protein encoded by the mRNA. (This effect is equivalent to that resulting from dominant-negative mutations, discussed in Chapter 9, although the cause is different.) Several mechanisms collectively termed **mRNA surveillance** help cells avoid the translation of improperly processed mRNA molecules. We have previously mentioned two such surveillance mechanisms: the recognition of improperly processed pre-mRNAs in the nucleus and their degradation by the exosome, and the general restriction against nuclear export of incompletely spliced pre-mRNAs that remain associated with a spliceosome.

Another mechanism called *nonsense-mediated decay* causes degradation of mRNAs in which one or more exons have been skipped during splicing. Except for pre-mRNAs that normally undergo alternative splicing, such exon skipping often will alter the open reading frame of the mRNA 3' to the improper exon junction, resulting in introduction of an incorrect stop codon. For nearly all properly spliced mRNAs, the stop codon is in the last exon. Nonsense-mediated decay results in the rapid degradation of mRNAs with stop codons that occur before the last splice junction in the mRNA.

A search for possible molecular signals that might indicate the positions of splice junctions in a processed mRNA led to the discovery of exon-junction complexes. As noted already, these complexes stimulate export of mRNPs from the nucleus. Analysis of yeast mutants suggests that some of the proteins in exon-junction complexes function in nonsense-mediated decay. One proposal based on these and other findings is that exon-junction complexes interact with a deadenylase that rapidly removes the poly(A) tail from an associated mRNA, leading to its rapid decapping and degradation by a 5' → 3' exonuclease (see Figure 12-29). In the case of properly spliced mRNAs, the exon-junction complexes are thought to be dislodged from the mRNA by passage of the first “pioneer” ribosome to translate the mRNA, thereby protecting the mRNA from degradation. For mRNAs with a stop codon before the final exon junction, however, one or more exon-junction complexes remain associated with the mRNA, resulting in nonsense-mediated decay.

Nonsense-mediated decay occurs in the cytoplasm of yeast cells. Remarkably, in mammalian cells, there is evi-

dence that the pioneer ribosome translates the mRNA while its 5' end is associated with the nuclear cap-binding complex and its poly(A) tail is associated with nuclear PABPII. This finding and other results raise the possibility that in cells of higher organisms, the first round of translation may occur in the nucleus as part of the nonsense-mediated decay mechanism of mRNA surveillance.

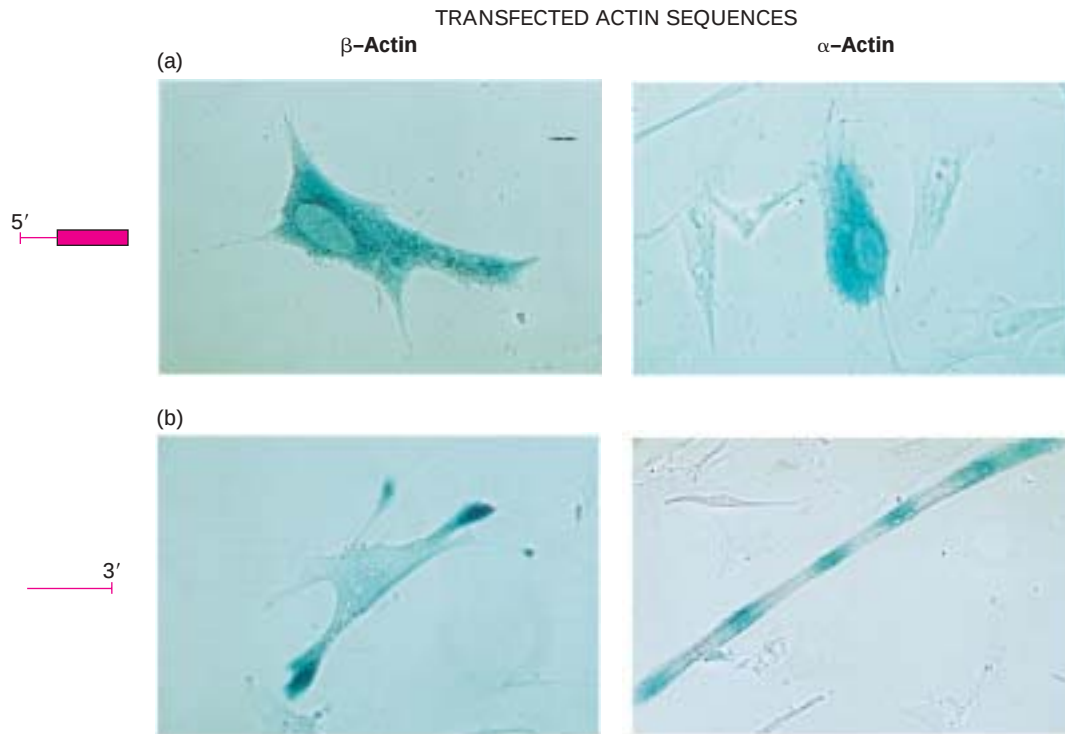
Localization of mRNAs Permits Production of Proteins at Specific Regions Within the Cytoplasm

Many cellular processes depend on localization of particular proteins to specific structures or regions of the cell. In later chapters we examine how some proteins are transported *after* their synthesis to their proper cellular location. Alternatively, protein localization might be achieved by localization of mRNAs to specific regions of the cell cytoplasm in which their encoded proteins function. In most cases examined thus far, such mRNA localization is specified by sequences in the 3' untranslated region of the mRNA.

A well-documented example of mRNA localization occurs in mammalian myoblasts (muscle precursor cells) as they differentiate into myotubes, the fused, multinucleated cells that make up muscle fibers. Myoblasts are motile cells that extend cytoplasmic regions, called *lamellipodia*, from the leading edge in the direction of movement. Extension of lamellipodia during cell movement requires polymerization of β -actin (Chapter 19). Sensibly, β -actin mRNA is concentrated in the leading edges of myoblasts, the region of the cell cytoplasm where the encoded protein is needed for motility. When myoblasts fuse into syncytial myotubes, β -actin expression is repressed and the muscle-specific α -actin is induced. In contrast to β -actin mRNA, α -actin mRNA is restricted to the perinuclear regions of myotubes. When cultured myoblasts in the process of differentiating are stained with fluorescent probes specific for α - or β -actin mRNA, both mRNAs are localized to their respective cellular regions.

To test the ability of actin mRNA sequences to direct the cytoplasmic localization of an mRNA, fragments of α - and β -actin cDNAs were inserted into separate plasmid vectors that express β -galactosidase from a strong viral promoter. The resulting plasmids then were transfected into cultured cells, which were assayed for β -galactosidase activity. These experiments showed that inclusion of the 3' untranslated end of α - or β -actin cDNAs directs localization of the expressed β -galactosidase, whereas the 5' untranslated and coding regions do not (Figure 12-31).

Treatment of cultured myoblasts with cytochalasin D, which disrupts actin microfilaments, leads to rapid delocalization of actin mRNAs, indicating that cytoskeletal actin microfilaments participate in the localization process. Disruption of other cytoskeletal components, however, does not alter the localization of actin mRNAs. Other types of evidence also implicate the actin cytoskeleton in mRNA localization. Presumably certain RNA-binding proteins interact



▲ EXPERIMENTAL FIGURE 12-31 Insertion of various sections of the actin genes in the 3' untranslated position of an expression vector showed that the 3' UTRs of α - and β -actin mRNAs direct localization of the reporter mRNA.

Plasmid expression vectors were constructed encoding β -galactosidase mRNAs containing as their 3' untranslated sequence either the 5' untranslated sequence and coding region of the α - or β -actin mRNAs (*top panels*), or the 3' untranslated sequence of these mRNAs (*bottom panels*), as indicated at the left. The recombinant plasmids were transfected separately into differentiating myoblasts. After a period of expression, the cells were fixed and then assayed for β -galactosidase activity by incubating them with X-gal, a substrate that is hydrolyzed by β -

galactosidase to yield a blue product. (a) Transfected cells that expressed engineered β -galactosidase mRNAs whose 3' untranslated region corresponded to the 5' untranslated sequence and coding region of α - or β -actin mRNA. These actin sequences did not cause localization of the β -galactosidase mRNA, as evidenced by the diffuse blue staining. (b) Transfected cells that expressed engineered β -galactosidase mRNAs whose 3' untranslated region corresponded to the 3' untranslated sequences of α - or β -actin mRNA. These sequences led to localization of β -galactosidase mRNA to lamellipodia in myoblasts (β -actin) or perinuclear regions in myotubes (α -actin), as evidenced by the deep staining in these regions. [Micrographs from E. H. Kislaukis et al., 1993, *J. Cell. Biol.* **123**:165.]

with specific sequences in the 3' untranslated regions of an mRNA and with specific components of microfilaments, including motor proteins that move cargo along the length of microfilaments. The best-understood example of this mechanism of mRNA localization occurs during cell division in *S. cerevisiae*, as we describe in Chapter 22.

KEY CONCEPTS OF SECTION 12.4

Cytoplasmic Mechanisms of Post-transcriptional Control

- Translation can be repressed by micro RNAs (miRNAs), which form imperfect hybrids with sequences in the 3' untranslated region (UTR) of specific target mRNAs.
- The related phenomenon of RNA interference, which probably evolved as an early defense system against viruses

and transposons, leads to degradation of mRNAs that form perfect hybrids with short interfering RNAs (siRNAs).

- Both miRNAs and siRNAs contain 21–23 nucleotides, are generated from longer precursor molecules, and are assembled into a multiprotein RNA-induced silencing complex (RISC) that either represses translation of target mRNAs or cleaves them (see Figure 12-27).
- Cytoplasmic polyadenylation is required for translation of mRNAs with a short poly(A) tail. Binding of a specific protein to regulatory elements in their 3' UTRs represses translation of these mRNAs. Phosphorylation of this RNA-binding protein, induced by an external signal, leads to lengthening of the 3' poly(A) tail and translation (see Figure 12-28).
- Most mRNAs are degraded as the result of the gradual shortening of their poly(A) tail (deadenylation) followed by

exosome-mediated 3' → 5' digestion or removal of the 5' cap and digestion by a 5' → 3' exonuclease (see Figure 12-29).

- Eukaryotic mRNAs encoding proteins that are expressed in short bursts generally have repeated copies of an AU-rich sequence in their 3' UTR. Specific proteins that bind to these elements also interact with the deadenylating enzyme and cytoplasmic exosomes, promoting rapid RNA degradation.

- Binding of various proteins to regulatory elements in the 3' or 5' UTRs of mRNAs regulates the translation or degradation of many mRNAs in the cytoplasm.

- Translation of ferritin mRNA and degradation of transferrin receptor (TfR) mRNA are both regulated by the same iron-sensitive RNA-binding protein. At low iron concentrations, this protein is in a conformation that binds to specific elements in the mRNAs, inhibiting ferritin mRNA translation or degradation of TfR mRNA (see Figure 12-30). This dual control precisely regulates the iron level within cells.

- Nonsense-mediated decay and other mRNA surveillance mechanisms prevent the translation of improperly processed mRNAs encoding abnormal proteins that might interfere with functioning of the corresponding normal proteins.

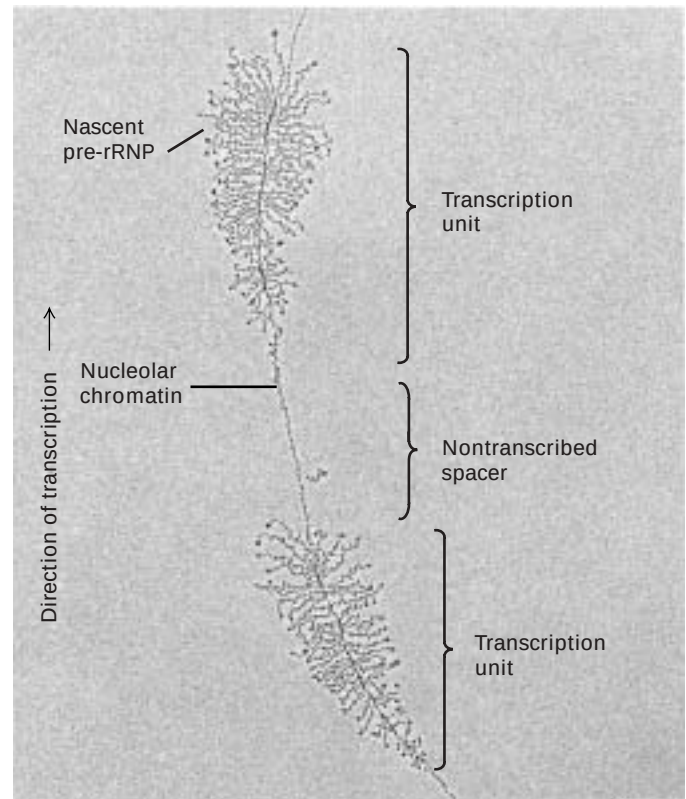
- Some mRNAs are directed to specific subcellular locations by sequences usually found in the 3' UTR, leading to localization of the encoded proteins.

12.5 Processing of rRNA and tRNA

Approximately 80 percent of the total RNA in rapidly growing mammalian cells (e.g., cultured HeLa cells) is rRNA, and 15 percent is tRNA; protein-coding mRNA thus constitutes only a small portion of the total RNA. The primary transcripts produced from most rRNA genes and from tRNA genes, like pre-mRNAs, are extensively processed to yield the mature, functional forms of these RNAs.

Pre-rRNA Genes Are Similar in All Eukaryotes and Function as Nucleolar Organizers

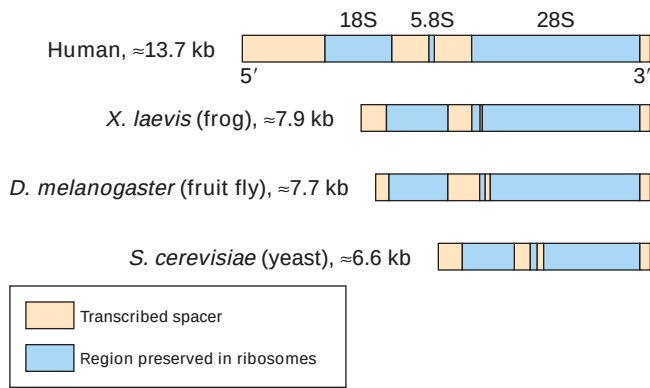
The 28S and 5.8S rRNAs associated with the large (60S) ribosomal subunit and the 18S rRNA associated with the small (40S) ribosomal subunit in higher eukaryotes (and the smaller, functionally equivalent rRNAs in all other eukaryotes) are encoded by a single type of **pre-rRNA** transcription unit. Transcription by RNA polymerase I yields a 45S primary transcript (pre-rRNA), which is processed into the mature 28S, 18S, and 5.8S rRNAs found in cytoplasmic ribosomes. Sequencing of the DNA encoding pre-rRNA from many species showed that this DNA shares several properties in all eukaryotes. First, the pre-rRNA genes are arranged in long tandem arrays separated by nontranscribed spacer regions ranging in length from ≈2 kb in frogs to ≈30 kb in hu-



▲ **FIGURE 12-32 Electron micrograph of pre-rRNA transcription units from nucleolus of a frog oocyte.** Each “feather” represents a pre-rRNA molecule associated with protein in a pre-ribonucleoprotein particle (pre-RNP) emerging from a transcription unit. Pre-rRNA transcription units are arranged in tandem, separated by nontranscribed spacer regions of nucleolar chromatin. [Courtesy of Y. Osheim and O. J. Miller, Jr.]

mans (Figure 12-32). Second, the genomic regions corresponding to the three mature rRNAs are always arranged in the same 5' → 3' order: 18S, 5.8S, and 28S. Third, in all eukaryotic cells (and even in bacteria), the pre-rRNA gene codes for, and the corresponding primary transcript contains, regions that are removed during processing and rapidly degraded. The general structure of pre-rRNAs is diagrammed in Figure 12-33.

Both the synthesis and processing of pre-rRNA occurs in the **nucleolus**. When pre-rRNA genes initially were identified in the nucleolus by in situ hybridization, it was not known whether any other DNA was required to form the nucleolus. Subsequent experiments with transgenic *Drosophila* strains demonstrated that a single complete pre-rRNA transcription unit induces formation of a small nucleolus. Thus a single pre-rRNA gene is sufficient to be a *nucleolar organizer*, and all the other components of the ribosome diffuse to the newly formed pre-rRNA. The structure of the nucleolus observed by light and electron microscopy results from the processing of pre-rRNA and the assembly of ribosomal subunits.



▲ **FIGURE 12-33 General structure of eukaryotic pre-rRNA transcription units.** The three coding regions (blue) encode the 18S, 5.8S, and 28S rRNAs found in ribosomes of higher eukaryotes or their equivalents in other species. The order of these coding regions in the genome is always 5' → 3'. Variations in the lengths of the transcribed spacer regions (tan) account for the major difference in the lengths of pre-rRNA transcription units in different organisms.

Small Nucleolar RNAs Assist in Processing Pre-rRNAs and Assembling Ribosome Subunits

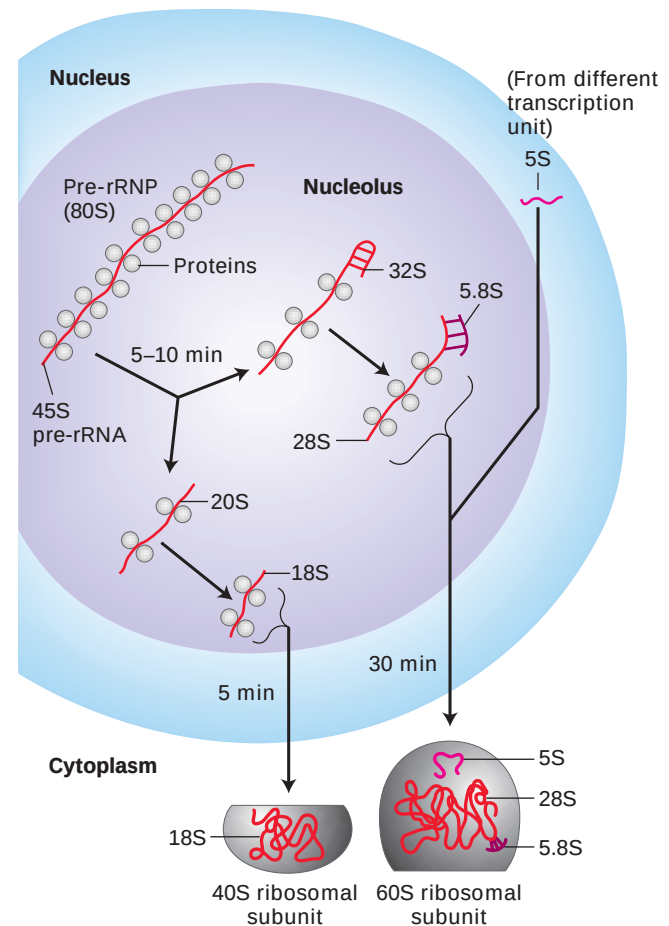
Following their synthesis in the nucleolus, nascent pre-rRNA transcripts are immediately bound by proteins, forming *pre-ribosomal ribonucleoprotein particles (pre-rRNPs)*. The largest of these (80S) contains an intact 45S pre-rRNA molecule, which is cut in a series of cleavage and exonucleolytic steps that ultimately yield the mature rRNAs found in ribosomes (Figure 12-34). During processing, pre-rRNA also is extensively modified, mostly by methylation of the 2'-hydroxyl group of specific riboses and conversion of specific uridine residues to pseudouridine. Some of the proteins in the pre-rRNPs found in nucleoli remain associated with the mature ribosomal subunits, whereas others are restricted to the nucleolus and assist in assembly of the subunits.

The positions of cleavage sites in pre-rRNA and the specific sites of 2'-O-methylation and pseudouridine formation are determined by approximately 150 different small nucleolus-restricted RNA species, called *small nucleolar RNAs (snoRNAs)*, which hybridize transiently to pre-rRNA molecules. Like the snRNAs that function in pre-mRNA processing, snoRNAs associate with proteins, forming ribonucleoprotein particles called *snoRNPs*. One large class of snoRNPs helps position a methyltransferase enzyme near methylation sites in the pre-mRNA. Another positions the enzyme that converts uridine to pseudouridine. Others function in the cleavage reactions that remove transcribed spacer regions. Once cleaved from pre-rRNAs, these sequences are degraded by the same exosome-associated 3' → 5' nuclear exonucleases that degrade introns spliced from pre-mRNAs.

Some snoRNAs are expressed from their own promoters by RNA polymerase II or III. Remarkably, however, the large

majority of snoRNAs are spliced-out introns of genes encoding functional mRNAs encoding proteins involved in ribosome synthesis or translation. Some snoRNAs are introns spliced from apparently nonfunctional mRNAs. The genes encoding these mRNAs seem to exist only to express snoRNAs from excised introns.

Unlike pre-rRNA genes, 5S rRNA genes are transcribed by RNA polymerase III in the nucleoplasm outside the nucleolus. Without further processing, 5S rRNA diffuses to the nucleolus, where it assembles with the 28S and 5.8S rRNAs and proteins into large ribosomal subunits (see Figure 12-34 and Figure 4-28). When assembly of ribosomal subunits in the nucleolus is complete, they are transported through nuclear pore complexes to the cytoplasm, where they appear first as free subunits. The ribosomal subunits are the largest cellular structures known to be transported through nuclear pore complexes.



▲ **FIGURE 12-34 Processing of pre-rRNA and assembly of ribosomes in higher eukaryotes.** Ribosomal and nucleolar proteins associate with 45S pre-rRNA as it is synthesized, forming an 80S pre-rRNP. Sites of cleavage and chemical modifications are determined by small nucleolar RNAs (not shown). Note that synthesis of 5S rRNA occurs outside the nucleolus.

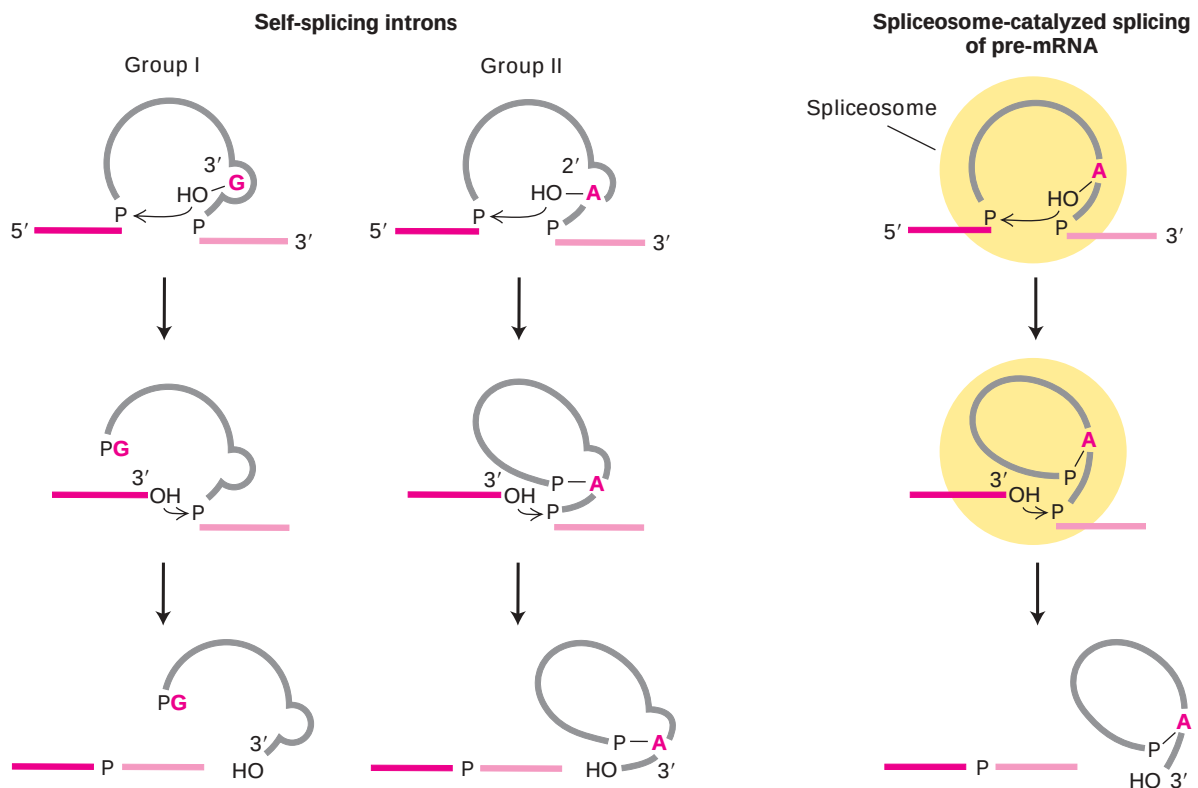
Self-Splicing Group I Introns Were the First Examples of Catalytic RNA

The DNA in the protozoan *Tetrahymena thermophila* contains an intervening intron in the region that encodes the large pre-rRNA molecule. Careful searches failed to uncover even one pre-rRNA gene without the extra sequence, indicating that splicing is required to produce mature rRNA in these organisms. In vitro studies showing that the pre-rRNA was spliced at the correct sites in the absence of any protein provided the first indication that RNA can function as a catalyst, like enzymes.

A whole raft of self-splicing sequences subsequently were found in pre-rRNAs from other single-celled organisms, in mitochondrial and chloroplast pre-rRNAs, in several pre-mRNAs from certain *E. coli* bacteriophages, and in some bacterial tRNA primary transcripts. The self-splicing sequences in all these precursors, referred to as *group I introns*, use guanosine as a cofactor and can fold by internal base pairing to juxtapose closely the two exons that must be joined. As discussed earlier, certain

mitochondrial and chloroplast pre-mRNAs and tRNAs contain a second type of self-splicing intron, designated *group II*.

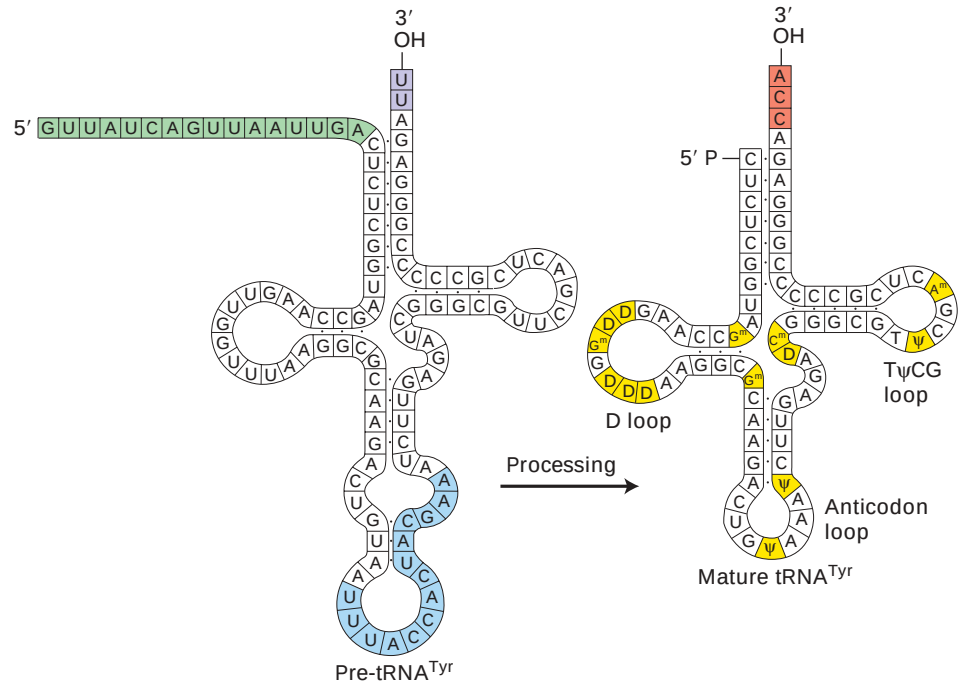
The splicing mechanisms used by group I introns, group II introns, and spliceosomes are generally similar, involving two transesterification reactions, which require no input of energy (Figure 12-35). Structural studies of the group I intron from *Tetrahymena* pre-rRNA combined with mutational and biochemical experiments have revealed that the RNA folds into a precise three-dimensional structure that, like protein enzymes, contains deep grooves for binding substrates and solvent-inaccessible regions that function in catalysis. The group I intron functions like a metalloenzyme to precisely orient the atoms that participate in the two transesterification reactions adjacent to catalytic Mg^{2+} ions. Considerable evidence now indicates that splicing by group II introns and by snRNAs in the spliceosome also involves bound catalytic Mg^{2+} ions. In both the groups I and II self-splicing introns and probably in the spliceosome, RNA functions as a **ribozyme**, an RNA sequence with catalytic ability.



▲ **FIGURE 12-35 Splicing mechanisms in group I and group II self-splicing introns and spliceosome-catalyzed splicing of pre-mRNA.** The intron is shown in gray, the exons to be joined in red. In group I introns, a guanosine cofactor (G) that is not part of the RNA chain associates with the active site. The 3'-hydroxyl group of this guanosine participates in a transesterification reaction with the phosphate at the 5' end of the intron; this

reaction is analogous to that involving the 2'-hydroxyl groups of the branch-site A in group II introns and pre-mRNA introns spliced in spliceosomes (see Figure 12-7). The subsequent transesterification that links the 5' and 3' exons is similar in all three splicing mechanisms. Note that spliced-out group I introns are linear structures, unlike the branched intron products in the other two cases. [Adapted from P. A. Sharp, 1987, *Science* **235**:769.]

► **FIGURE 12-36 Changes that occur during processing of tyrosine pre-tRNA.** A 14-nucleotide intron (blue) in the anticodon loop is removed by splicing. A 16-nucleotide sequence (green) at the 5' end is cleaved by RNase P. U residues at the 3' end are replaced by the CCA sequence (red) found in all mature tRNAs. Numerous bases in the stem-loops are converted to characteristic modified bases (yellow). Not all pre-tRNAs contain introns that are spliced out during processing, but they all undergo the other types of changes shown here. D = dihydrouridine; Ψ = pseudouridine.



Pre-tRNAs Undergo Cleavage, Base Modification, and Sometimes Protein-Catalyzed Splicing

Mature cytosolic tRNAs, which average 75–80 nucleotides in length, are produced from larger precursors (pre-tRNAs) synthesized by RNA polymerase III in the nucleoplasm. Mature tRNAs also contain numerous modified bases that are not present in tRNA primary transcripts. Cleavage and base modification occur during processing of all pre-tRNAs; some pre-tRNAs also are spliced during processing.

A 5' sequence of variable length that is absent from mature tRNAs is present in all pre-tRNAs (Figure 12-36). These extra 5' nucleotides are removed by *ribonuclease P* (*RNase P*), a ribonucleoprotein endonuclease. Studies with *E. coli* RNase P indicate that at high Mg^{2+} concentrations, the RNA component alone can recognize and cleave *E. coli* pre-tRNAs. The RNase P polypeptide increases the rate of cleavage by the RNA, allowing it to proceed at physiological Mg^{2+} concentrations. A comparable RNase P functions in eukaryotes.

About 10 percent of the bases in pre-tRNAs are modified enzymatically during processing. Three classes of base modifications occur (see Figure 12-36): replacement of U residues at the 3' end of pre-tRNA with a CCA sequence, which is found at the 3' end of all tRNAs and is required for their charging by aminoacyl-tRNA synthetases during protein synthesis; addition of methyl and isopentenyl groups to the heterocyclic ring of purine bases and methylation of the 2'-OH group in the ribose of any residue; and conversion of specific uridines to dihydrouridine, pseudouridine, or ribothymidine residues.

As shown in Figure 12-36, the pre-tRNA expressed from the yeast tyrosine tRNA ($tRNA^{Tyr}$) gene contains a

14-base intron that is not present in mature $tRNA^{Tyr}$. Some other eukaryotic tRNA genes and some archaeal tRNA genes also contain introns. The introns in nuclear pre-tRNAs are shorter than those in pre-mRNAs and lack the consensus splice-site sequences found in pre-mRNAs (see Figure 12-6). Pre-tRNA introns also are clearly distinct from the much longer self-splicing group I and group II introns found in chloroplast and mitochondrial pre-rRNAs. The mechanism of pre-tRNA splicing differs in three fundamental ways from the mechanisms utilized by self-splicing introns and spliceosomes (see Figure 12-35). First, splicing of pre-tRNAs is catalyzed by proteins, not by RNAs. Second, a pre-tRNA intron is excised in one step that entails simultaneous cleavage at both ends of the intron. Finally, hydrolysis of GTP and ATP is required to join the two tRNA halves generated by cleavage on either side of the intron.

After pre-tRNAs are processed in the nucleoplasm, the mature tRNAs are transported to the cytoplasm through nuclear pore complexes by exportin-t, as discussed previously. In the cytoplasm, tRNAs are passed between aminoacyl-tRNA synthetases, elongation factors, and ribosomes during protein synthesis (Chapter 4). Thus tRNAs generally are associated with proteins and spend little time free in the cell, as is also the case for mRNAs and rRNAs.

KEY CONCEPTS OF SECTION 12.5

Processing of rRNA and tRNA

- A large precursor pre-rRNA (45S in humans) synthesized by RNA polymerase I undergoes cleavage, exonucle-

olytic digestion, and base modifications to yield mature 28S, 18S, and 5.8S rRNAs, which associate with ribosomal proteins into ribosomal subunits (see Figure 12-34).

- Synthesis and processing of pre-rRNA occur in the nucleolus. The 5S rRNA component of the large ribosomal subunit is synthesized in the nucleoplasm by RNA polymerase III and is not processed.
- Group I and group II self-splicing introns and snRNAs in spliceosomes all function as ribozymes, or catalytically active RNA sequences, that carry out splicing by analogous transesterification reactions requiring bound Mg^{+2} ions (see Figure 12-35).
- Pre-tRNAs synthesized by RNA polymerase III in the nucleoplasm are processed by removal of the 5'-end sequence, addition of CCA to the 3' end, and modification of multiple internal bases (see Figure 12-36).
- Some pre-tRNAs contain a short intron that is removed by a protein-catalyzed mechanism distinct from the splicing of pre-mRNA and self-splicing introns.
- All species of RNA molecules are associated with proteins in various types of ribonucleoprotein particles both in the nucleus and after export to the cytoplasm.

PERSPECTIVES FOR THE FUTURE

In this and the previous chapter, we have seen that in eukaryotic cells, mRNAs are synthesized and processed in the nucleus, transported through nuclear pore complexes to the cytoplasm, and then, in some cases, transported to specific areas of the cytoplasm before being translated by ribosomes. Each of these fundamental processes is carried out by complex macromolecular machines composed of scores of proteins and in many cases RNAs, as well. The complexity of these macromolecular machines ensures accuracy in finding promoters and splice sites in the long length of DNA and RNA sequences and provides various avenues for regulating synthesis of a polypeptide chain. Much remains to be learned about the structure, operation, and regulation of such complex machines as spliceosomes and the cleavage/polyadenylation apparatus.

Recent examples of the regulation of pre-mRNA splicing raise the question of how extracellular signals might control such events, especially in the nervous system of vertebrates. A case in point is the remarkable situation in the chick inner ear where multiple isoforms of the Ca^{2+} -activated K^+ channel called *Slo* are produced by alternative RNA splicing. Cell-cell interactions appear to inform cells of their position in the cochlea, leading to alternative splicing of *Slo* pre-mRNA. The challenging task facing researchers is to discover how such cell-cell interactions regulate the activity of RNA-processing factors.

The mechanism of mRNP transport through nuclear pore complexes poses many intriguing questions. Future research will likely reveal additional activities of hnRNP and nuclear mRNP proteins and clarify their mechanisms of action. For instance, there is a small gene family encoding proteins homologous to the large subunit of the mRNA-exporter. What are the functions of these related proteins? Do they participate in the transport of overlapping sets of mRNPs? Some hnRNP proteins contain nuclear-retention signals that prevent nuclear export when fused to hnRNP proteins with nuclear-export signals (NESs). How are these hnRNP proteins selectively removed from processed mRNAs in the nucleus, allowing the mRNAs to be transported to the cytoplasm?

The localization of certain mRNAs to specific subcellular locations is fundamental to the development of multicellular organisms. As discussed in Chapter 22, during development an individual cell frequently divides into daughter cells that function differently from each other. In the language of developmental biology, the two daughter cells are said to have different developmental fates. In many cases, this difference in developmental fate results from the localization of an mRNA to one region of the cell before mitosis so that after cell division, it is present in one daughter cell and not the other. Much exciting work remains to be done to fully understand the molecular mechanisms controlling mRNA localization that are critical for the normal development of multicellular organisms.

Some of the most exciting and unanticipated discoveries in molecular cell biology in recent years concern the existence and function of miRNAs and the process of RNA interference. RNA interference (RNAi) provides molecular cell biologists with a powerful method for studying gene function. The discovery of more than a hundred miRNAs in humans and other organisms suggests that multiple significant examples of translational control by this mechanism await to be characterized. Recent studies in plants link similar short nuclear RNAs to the control of DNA methylation and the formation of heterochromatin. Will similar processes control gene expression through the assembly of heterochromatin in humans and other animals? What other regulatory processes might be directed by micro RNAs? Since control by these mechanisms depends on base pairing between miRNAs and target mRNAs or genes, genomic and bioinformatic methods will probably suggest genes that may be controlled by these mechanisms. What other processes in addition to translation control, mRNA degradation, and heterochromatin assembly might be controlled by miRNAs?

These are just a few of the fascinating questions concerning RNA processing, post-transcriptional control, and nuclear transport that will challenge molecular cell biologists in the coming decades. The astounding discoveries of entirely unanticipated mechanisms of gene control by miRNAs remind us that there will likely be many more surprises in the future.

KEY TERMS

5' cap 493	mRNA-exporter 514
alternative splicing 505	mRNA surveillance 523
cargo complex 510	nuclear pore complex (NPC) 509
cleavage/polyadenylation complex 497	poly(A) tail 496
cross-exon recognition complex 501	pre-mRNA 494
exosome 504	pre-rRNA 525
exportin 1 512	Ran protein 510
FG-nucleoporins 510	ribozyme 527
group I introns 502	RNA editing 508
group II introns 502	RNA recognition motif (RRM) 495
importins 510	RNA splicing 497
iron-response element-binding protein (IRE-BP) 522	RNA-induced silencing complex (RISC) 518
karyopherins 513	small nuclear RNAs (snRNAs) 499
micro RNAs (miRNAs) 518	spliceosome 499
	SR proteins 501

REVIEW THE CONCEPTS

1. Describe three types of post-transcriptional regulation of protein-coding genes.
2. What is the evidence that transcription termination by RNA polymerase II is coupled to polyadenylation?
3. It has been suggested that manipulation of HIV anti-termination might provide for effective therapies in combating AIDS. What effect would a mutation in the TAR sequence that abolishes Tat binding have on HIV transcription after HIV infection and why? A mutation in Cdk9 that abolishes activity?
4. Some heat shock genes encode proteins that act rapidly to protect other proteins from harsh conditions. Describe the mechanism that has evolved to regulate the expression of such genes.
5. You are investigating the transcriptional regulation of genes from a recently discovered eukaryotic organism. You find that this organism has three RNA polymerase-like enzymes and that genes encoding proteins are transcribed using an RNA polymerase II-like enzyme. When examining the full-length mRNAs produced, you find that there is no 5' structure (i.e., cap). Are you surprised? Why/why not?
6. Why do researchers believe that the fragile-X disorder involves an aberration in RNA binding?
7. What is the difference between hnRNAs, snRNAs, miRNAs, siRNAs, and snoRNAs?

8. Describe how the discovery that introns are removed during splicing was made. How are the locations of exon-intron junctions predicted?

9. What are the mechanistic similarities between Group II intron self-splicing and spliceosomal splicing? What is the evidence that there may be an evolutionary relationship between the two?

10. Where do researchers believe most transcription and RNA-processing events occur? What is the evidence to support this?

11. You obtain the sequence of a gene containing 10 exons, 9 introns, and a 3' UTR containing a polyadenylation consensus sequence. The fifth intron also contains a polyadenylation site. To test whether both polyadenylation sites are used, you isolate mRNA and find a longer transcript from muscle tissue and a shorter mRNA transcript from all other tissues. Speculate about the mechanism involved in the production of these two transcripts.

12. What is the evidence that an mRNA exporter directs mRNPs through nuclear pores? What is the evidence that most vertebrates also utilize such an exporter?

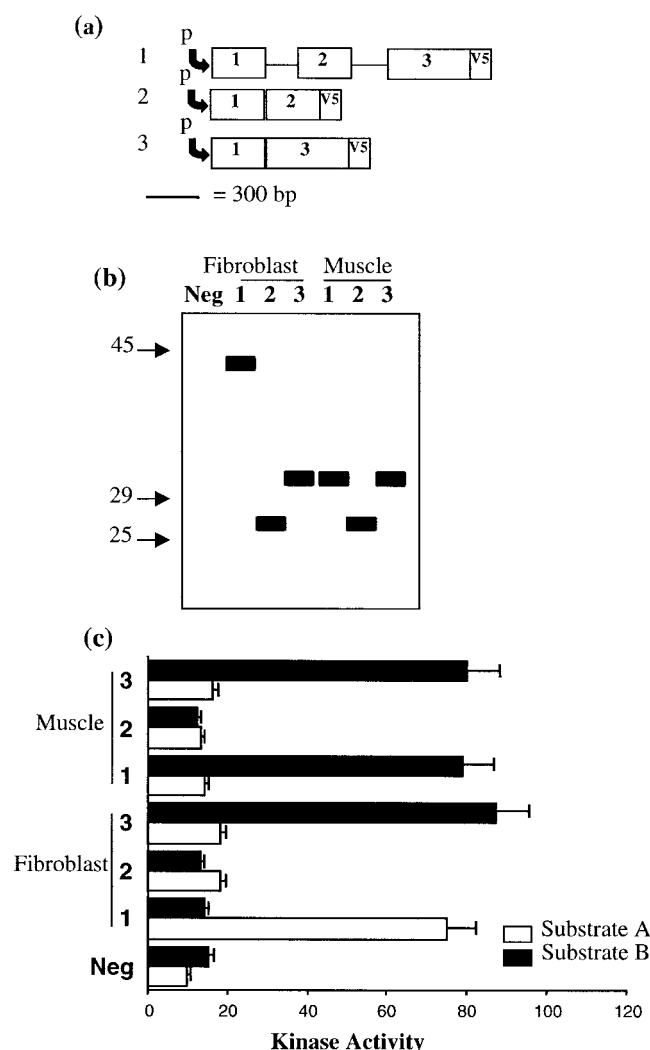
13. Why is localization of Ran GAP in the nucleus and Ran nucleotide exchange factor (RCC1) in the cytoplasm necessary for unidirectional transport of cargo proteins containing an NES?

14. Speculate about why plants deficient in Dicer activity show increased sensitivity to infection by RNA viruses.

15. What is the evidence that some mRNAs are directed to accumulate in specific subcellular locations?

ANALYZE THE DATA

A specific protein kinase (PK) gene is speculated to be differentially spliced in muscle tissue. This gene comprises three exons and two intron sequences and in fibroblast cells encodes a 38.5-kD protein. Investigators have transfected various portions of a genomic or cDNA copy of the PK gene into both muscle and fibroblast cells (see the constructs in part (a) of the figure). The expression system utilizes a promoter active in both cell types and a C-terminal fusion to a small epitope tag called V5, which contributes ~5.5 kD to the fusion protein. Part (b) of the figure shows the results of an immunoprecipitation experiment designed to analyze the expression products of the transfected cells. A negative control (Neg) of untransfected muscle cells is included. Molecular weight markers in kD are indicated on the left. Immunoprecipitated proteins as shown in part (b) were then placed in protein kinase assays with two different substrates, A and B, to discern activity of the expressed proteins.



- What can be concluded about differential splicing of the PK gene in fibroblast versus muscle cells? Are there other experiments that could confirm these results?
- What sequence or sequences contribute to regulation of alternative splicing?
- How does the presence or absence of exon 3 alter the catalytic activity of the encoded PK protein? Which data support your conclusions?

REFERENCES

Processing of Eukaryotic Pre-mRNA

- Bentley, D. 2002. The mRNA assembly line: transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* **14**:336–342.
- Bonen, L., and J. Vogel. 2001. The ins and outs of group II introns. *Trends Genet.* **17**:322–331.
- Butler, J. S. 2002. The yin and yang of the exosome. *Trends Cell Biol.* **12**:90–96.

Dreyfuss, G., V. N. Kim, and N. Kataoka. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell Biol.* **3**:195–205.

Moore, M. J. 2002. Nuclear RNA turnover. *Cell* **108**:431–434.

Shatkin, A. J., and J. L. Manley. 2000. The ends of the affair: capping and polyadenylation. *Nature Struct. Biol.* **7**:838–842.

Villa, T., J. A. Pleiss, and C. Guthrie. 2002. Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell* **109**:149–152.

Regulation of Pre-mRNA Processing

Black, D. L. 2003. Mechanisms of alternative pre-mRNA splicing. *Ann. Rev. Biochem.* **72**:291–336.

Blanc, V., and N. O. Davidson. 2003. C-to-U RNA editing: mechanisms leading to genetic diversity. *J. Biol. Chem.* **278**:1395–1398.

Faustino, N. A., and T. A. Cooper. 2003. Pre-mRNA splicing and human disease. *Genes Devel.* **17**:419–437.

Maas, S., A. Rich, and K. Nishikura. 2003. A-to-I RNA editing: recent news and residual mysteries. *J. Biol. Chem.* **278**:1391–1394.

Maniatis, T., and B. Tasic. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**:236–243.

Macromolecular Transport Across the Nuclear Envelope

Chook, Y. M., and G. Blobel. 2001. Karyopherins and nuclear import. *Curr. Opin. Struct. Biol.* **11**:703–715.

Conti, E., and E. Izaurralde. 2001. Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.* **13**:310–319.

Johnson, A. W., E. Lund, and J. Dahlberg. 2002. Nuclear export of ribosomal subunits. *Trends Biochem. Sci.* **27**:580–585.

Reed, R., and E. Hurt. 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**:523–531.

Ribbeck, K., and D. Gorlich. 2001. Kinetic analysis of translocation through nuclear pore complexes. *Embo. J.* **20**:1320–1330.

Rout, M. P., and J. D. Aitchison. 2001. The nuclear pore complex as a transport machine. *J. Biol. Chem.* **276**:16593–16596.

Cytoplasmic Mechanisms of Post-transcriptional Control

Ambros, V. 2001. MicroRNAs: tiny regulators with great potential. *Cell* **107**:823–826.

Cerutti, H. 2003. RNA interference: traveling in the cell and gaining functions? *Trends Genet.* **19**:39–46.

Hannon, G. J. 2002. RNA interference. *Nature* **418**:244–251.

Kloc, M., N. R. Zearfoss, and L. D. Etkin. 2002. Mechanisms of subcellular mRNA localization. *Cell* **108**:533–544.

Maquat, L. E., and G. G. Carmichael. 2001. Quality control of mRNA function. *Cell* **104**:173–176.

Mendez, R., and J. D. Richter. 2001. Translational control by CPEB: a means to the end. *Nat. Rev. Mol. Cell Biol.* **2**:521–529.

Van Hoof, A., and R. Parker. 2002. Messenger RNA degradation: beginning at the end. *Curr. Biol.* **12**:R285–R287.

Wilkinson, M. F., and A. B. Shyu. 2002. RNA surveillance by nuclear scanning? *Nature Cell Biol.* **4**:E144–E147.

Processing of rRNA and tRNA

Fatica, A., and D. Tollervey. 2002. Making ribosomes. *Curr. Opin. Cell Biol.* **14**:313–318.

Hopper, A. K., and E. M. Phizicky. 2003. tRNA transfers to the limelight. *Genes Devel.* **17**:162–180.