



chapter 26

RNA METABOLISM

26.1 DNA-Dependent Synthesis of RNA 996

26.2 RNA Processing 1007

26.3 RNA-Dependent Synthesis of RNA and DNA 1021

The RNA of the cell is partly in the nucleus, partly in particles in the cytoplasm and partly as the “soluble” RNA of the cell sap; many workers have shown that all these three fractions turn over differently. It is very important to realize in any discussion of the role of RNA in the cell that it is very inhomogeneous metabolically, and probably of more than one type.

—Francis H. C. Crick, *article in Symposia of the Society for Experimental Biology*, 1958

Expression of the information in a gene generally involves production of an RNA molecule transcribed from a DNA template. Strands of RNA and DNA may seem quite similar at first glance, differing only in that RNA has a hydroxyl group at the 2' position of the aldopentose and uracil instead of thymine. However, unlike DNA, most RNAs carry out their functions as single strands, strands that fold back on themselves and have the potential for much greater structural diversity than DNA (Chapter 8). RNA is thus suited to a variety of cellular functions.

RNA is the only macromolecule known to have a role both in the storage and transmission of information and in catalysis, which has led to much speculation about its possible role as an essential chemical intermediate in the development of life on this planet. The discovery of catalytic RNAs, or ribozymes, has changed the very definition of an enzyme, extending it beyond the domain of proteins. Proteins nevertheless remain essential to RNA and its functions. In the modern cell, all nucleic acids, including RNAs, are complexed with proteins. Some of these complexes are quite elaborate, and

RNA can assume both structural and catalytic roles within complicated biochemical machines.

All RNA molecules except the RNA genomes of certain viruses are derived from information permanently stored in DNA. During **transcription**, an enzyme system converts the genetic information in a segment of double-stranded DNA into an RNA strand with a base sequence complementary to one of the DNA strands. Three major kinds of RNA are produced. **Messenger RNAs (mRNAs)** encode the amino acid sequence of one or more polypeptides specified by a gene or set of genes. **Transfer RNAs (tRNAs)** read the information encoded in the mRNA and transfer the appropriate amino acid to a growing polypeptide chain during protein synthesis. **Ribosomal RNAs (rRNAs)** are constituents of ribosomes, the intricate cellular machines that synthesize proteins. Many additional specialized RNAs have regulatory or catalytic functions or are precursors to the three main classes of RNA.

During replication the entire chromosome is usually copied, but transcription is more selective. Only particular genes or groups of genes are transcribed at any one time, and some portions of the DNA genome are never transcribed. The cell restricts the expression of genetic information to the formation of gene products needed at any particular moment. Specific regulatory sequences mark the beginning and end of the DNA segments to be transcribed and designate which strand in duplex DNA is to be used as the template. The regulation of transcription is described in detail in Chapter 28.

In this chapter we examine the synthesis of RNA on a DNA template and the postsynthetic processing and turnover of RNA molecules. In doing so we encounter many of the specialized functions of RNA, including catalytic functions. Interestingly, the substrates for RNA enzymes are often other RNA molecules. We also describe systems in which RNA is the template and DNA the product, rather than vice versa. The information pathways thus come full circle, revealing that template-dependent nucleic acid synthesis has standard rules

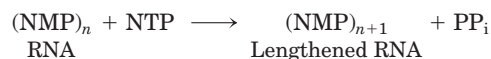
regardless of the nature of template or product (RNA or DNA). This examination of the biological interconversion of DNA and RNA as information carriers leads to a discussion of the evolutionary origin of biological information.

26.1 DNA-Dependent Synthesis of RNA

Our discussion of RNA synthesis begins with a comparison between transcription and DNA replication (Chapter 25). Transcription resembles replication in its fundamental chemical mechanism, its polarity (direction of synthesis), and its use of a template. And like replication, transcription has initiation, elongation, and termination phases—though in the literature on transcription, initiation is further divided into discrete phases of DNA binding and initiation of RNA synthesis. Transcription differs from replication in that it does not require a primer and, generally, involves only limited segments of a DNA molecule. Additionally, within transcribed segments only one DNA strand serves as a template.

RNA Is Synthesized by RNA Polymerases

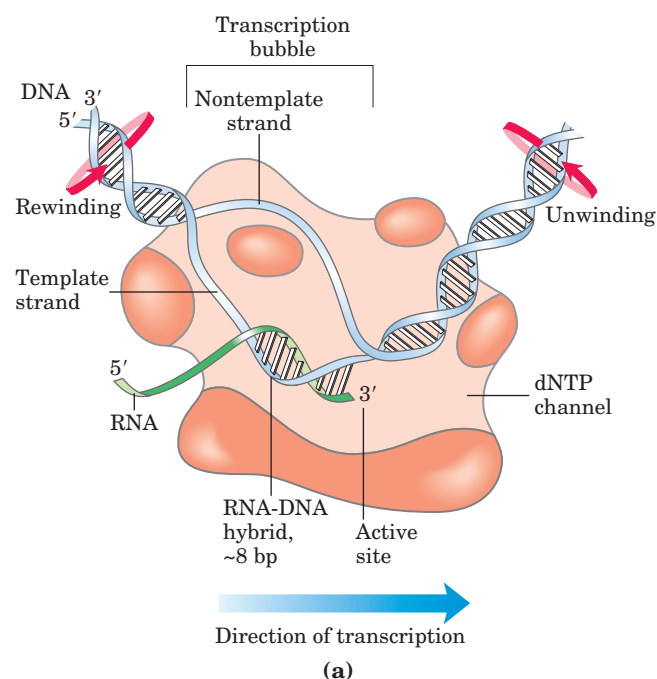
The discovery of DNA polymerase and its dependence on a DNA template spurred a search for an enzyme that synthesizes RNA complementary to a DNA strand. By 1960, four research groups had independently detected an enzyme in cellular extracts that could form an RNA polymer from ribonucleoside 5'-triphosphates. Subsequent work on the purified *Escherichia coli* RNA polymerase helped to define the fundamental properties of transcription (Fig. 26-1). **DNA-dependent RNA polymerase** requires, in addition to a DNA template, all four ribonucleoside 5'-triphosphates (ATP, GTP, UTP, and CTP) as precursors of the nucleotide units of RNA, as well as Mg^{2+} . The protein also binds one Zn^{2+} . The chemistry and mechanism of RNA synthesis closely resemble those used by DNA polymerases (see Fig. 25-5). RNA polymerase elongates an RNA strand by adding ribonucleotide units to the 3'-hydroxyl end, building RNA in the 5'→3' direction. The 3'-hydroxyl group acts as a nucleophile, attacking the α phosphate of the incoming ribonucleoside triphosphate (Fig. 26-1b) and releasing pyrophosphate. The overall reaction is



RNA polymerase requires DNA for activity and is most active when bound to a double-stranded DNA. As noted above, only one of the two DNA strands serves as a template. The template DNA strand is copied in the 3'→5' direction (antiparallel to the new RNA strand), just as in DNA replication. Each nucleotide in the newly formed RNA is selected by Watson-Crick base-pairing interac-

tions; U residues are inserted in the RNA to pair with A residues in the DNA template, G residues are inserted to pair with C residues, and so on. Base-pair geometry (see Fig. 25-6) may also play a role in base selection.

Unlike DNA polymerase, RNA polymerase does not require a primer to initiate synthesis. Initiation occurs when RNA polymerase binds at specific DNA sequences called promoters (described below). The 5'-triphosphate group of the first residue in a nascent (newly formed) RNA molecule is not cleaved to release PP_i , but instead remains intact throughout the transcription process. During the elongation phase of transcription, the growing end of the new RNA strand base-pairs temporarily with the DNA template to form a short hybrid



MECHANISM FIGURE 26-1 Transcription by RNA polymerase in *E. coli*. For synthesis of an RNA strand complementary to one of two DNA strands in a double helix, the DNA is transiently unwound. (a) About 17 bp are unwound at any given time. RNA polymerase and the bound transcription bubble move from left to right along the DNA as shown; facilitating RNA synthesis. The DNA is unwound ahead and rewound behind as RNA is transcribed. Red arrows show the direction in which the DNA must rotate to permit this process. As the DNA is rewound, the RNA-DNA hybrid is displaced and the RNA strand extruded. The RNA polymerase is in close contact with the DNA ahead of the transcription bubble, as well as with the separated DNA strands and the RNA within and immediately behind the bubble. A channel in the protein funnels new nucleoside triphosphates (NTPs) to the polymerase active site. The polymerase footprint encompasses about 35 bp of DNA during elongation.

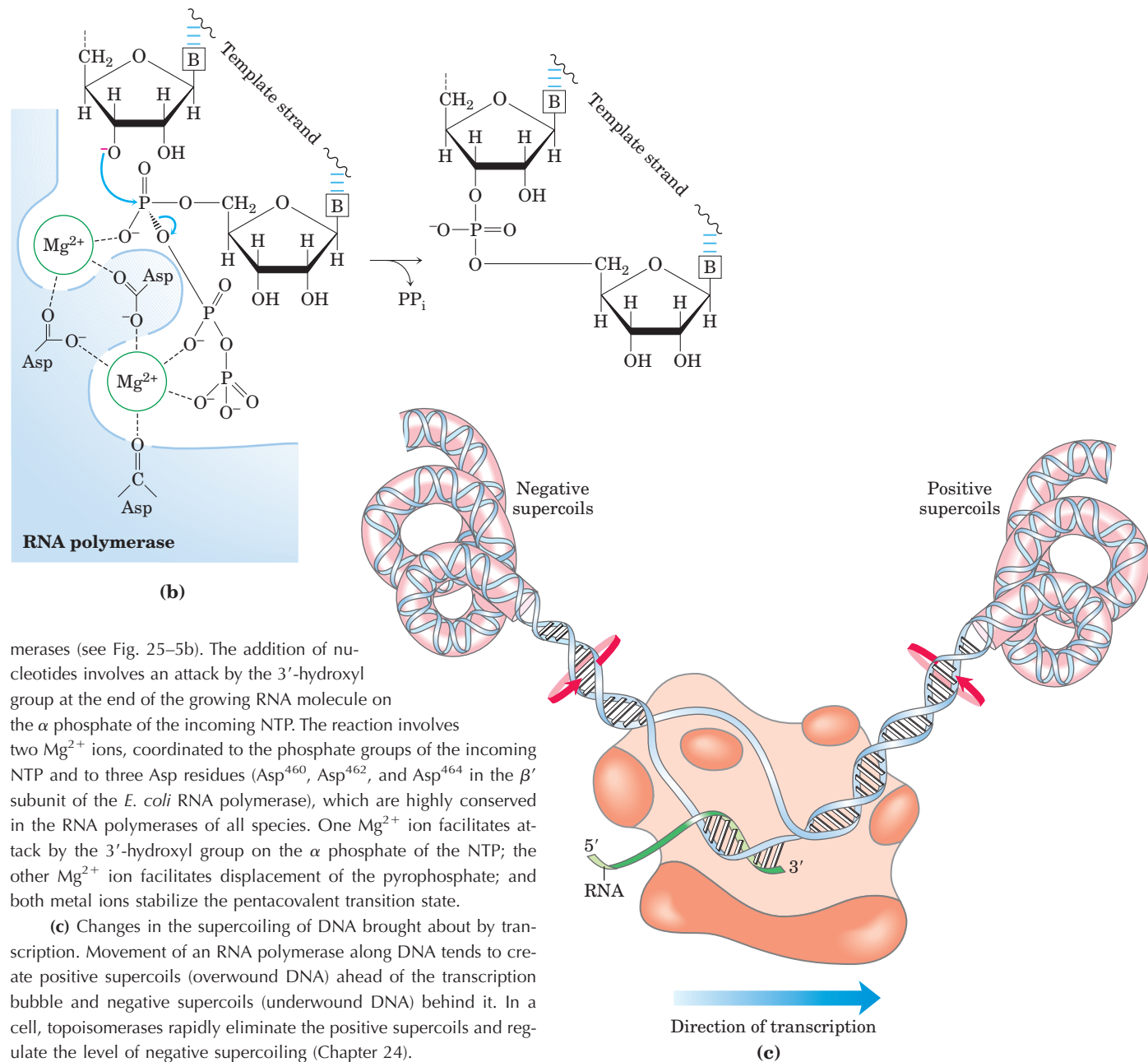
(b) Catalytic mechanism of RNA synthesis by RNA polymerase. Note that this is essentially the same mechanism used by DNA poly-

RNA-DNA double helix, estimated to be 8 bp long (Fig. 26-1a). The RNA in this hybrid duplex “peels off” shortly after its formation, and the DNA duplex re-forms.

To enable RNA polymerase to synthesize an RNA strand complementary to one of the DNA strands, the DNA duplex must unwind over a short distance, forming a transcription “bubble.” During transcription, the *E. coli* RNA polymerase generally keeps about 17 bp unwound. The 8 bp RNA-DNA hybrid occurs in this unwound region. Elongation of a transcript by *E. coli* RNA polymerase proceeds at a rate of 50 to 90 nucleotides/s. Because DNA is a helix, movement of a transcription bubble requires considerable strand rotation of the nucleic acid molecules. DNA strand rotation is restricted

in most DNAs by DNA-binding proteins and other structural barriers. As a result, a moving RNA polymerase generates waves of positive supercoils ahead of the transcription bubble and negative supercoils behind (Fig. 26-1c). This has been observed both in vitro and in vivo (in bacteria). In the cell, the topological problems caused by transcription are relieved through the action of topoisomerases (Chapter 24).

The two complementary DNA strands have different roles in transcription. The strand that serves as template for RNA synthesis is called the **template strand**. The DNA strand complementary to the template, the **nontemplate strand**, or **coding strand**, is identical in base sequence to the RNA transcribed from the gene,



merases (see Fig. 25-5b). The addition of nucleotides involves an attack by the 3'-hydroxyl group at the end of the growing RNA molecule on the α phosphate of the incoming NTP. The reaction involves two Mg^{2+} ions, coordinated to the phosphate groups of the incoming NTP and to three Asp residues (Asp⁴⁶⁰, Asp⁴⁶², and Asp⁴⁶⁴ in the β' subunit of the *E. coli* RNA polymerase), which are highly conserved in the RNA polymerases of all species. One Mg^{2+} ion facilitates attack by the 3'-hydroxyl group on the α phosphate of the NTP; the other Mg^{2+} ion facilitates displacement of the pyrophosphate; and both metal ions stabilize the pentacoordinate transition state.

(c) Changes in the supercoiling of DNA brought about by transcription. Movement of an RNA polymerase along DNA tends to create positive supercoils (overwound DNA) ahead of the transcription bubble and negative supercoils (underwound DNA) behind it. In a cell, topoisomerases rapidly eliminate the positive supercoils and regulate the level of negative supercoiling (Chapter 24).

(5') CGCTATAGCGTTT(3')	DNA nontemplate (coding) strand
(3') GCGATATCGCAA(5')	DNA template strand
(5') CGCUAUAGCGUUU(3')	RNA transcript

with U in the RNA in place of T in the DNA (Fig. 26-2). The coding strand for a particular gene may be located in either strand of a given chromosome (as shown in Fig. 26-3 for a virus). The regulatory sequences that control transcription (described later in this chapter) are by convention designated by the sequences in the coding strand.

The DNA-dependent RNA polymerase of *E. coli* is a large, complex enzyme with five core subunits ($\alpha_2\beta\beta'\omega$; M_r 390,000) and a sixth subunit, one of a group designated σ , with variants designated by size (molecular weight). The σ subunit binds transiently to the core and directs the enzyme to specific binding sites on the DNA (described below). These six subunits constitute the RNA polymerase holoenzyme (Fig. 26-4). The RNA polymerase holoenzyme of *E. coli* thus exists in several forms, depending on the type of σ subunit. The most common subunit is σ^{70} (M_r 70,000), and the upcoming discussion focuses on the corresponding RNA polymerase holoenzyme.

RNA polymerases lack a separate proofreading 3'→5' exonuclease active site (such as that of many DNA polymerases), and the error rate for transcription is higher than that for chromosomal DNA replication—approximately one error for every 10^4 to 10^5 ribonucleotides incorporated into RNA. Because many copies of an RNA are generally produced from a single gene and all RNAs are eventually degraded and replaced, a mistake in an RNA molecule is of less consequence to the cell than a mistake in the permanent information stored in DNA. Many RNA polymerases, including bacterial RNA polymerase and the eukaryotic RNA polymerase II (discussed below), do pause when a mispaired base is added during transcription, and they can remove

FIGURE 26-2 Template and nontemplate (coding) DNA

strands. The two complementary strands of DNA are defined by their function in transcription. The RNA transcript is synthesized on the template strand and is identical in sequence (with U in place of T) to the nontemplate strand, or coding strand.

mismatched nucleotides from the 3' end of a transcript by direct reversal of the polymerase reaction. But we do not yet know whether this activity is a true proofreading function and to what extent it may contribute to the fidelity of transcription.

RNA Synthesis Begins at Promoters

Initiation of RNA synthesis at random points in a DNA molecule would be an extraordinarily wasteful process. Instead, an RNA polymerase binds to specific sequences in the DNA called **promoters**, which direct the transcription of adjacent segments of DNA (genes). The sequences where RNA polymerases bind can be quite variable, and much research has focused on identifying the particular sequences that are critical to promoter function.

In *E. coli*, RNA polymerase binding occurs within a region stretching from about 70 bp before the transcription start site to about 30 bp beyond it. By convention, the DNA base pairs that correspond to the beginning of an RNA molecule are given positive numbers, and those preceding the RNA start site are given negative numbers. The promoter region thus extends between positions -70 and +30. Analyses and comparisons of the most common class of bacterial promoters (those recognized by an RNA polymerase holoenzyme containing σ^{70}) have revealed similarities in two short sequences centered about positions -10 and -35 (Fig. 26-5). These sequences are important interaction sites for the σ^{70} subunit. Although the sequences are not identical for all bacterial promoters in this class, certain nucleotides that are particularly common at each position form a **consensus sequence** (recall the *E. coli*

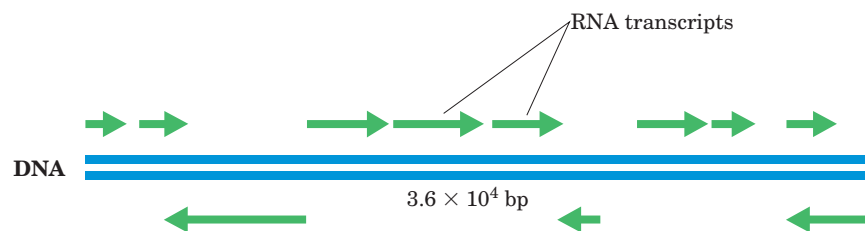


FIGURE 26-3 Organization of coding information in the adenovirus genome. The genetic information of the adenovirus genome (a conveniently simple example) is encoded by a double-stranded DNA molecule of 36,000 bp, both strands of which encode proteins. The information for most proteins is encoded by the top strand—but by convention, the strand transcribed from left to right—but some is encoded by the bottom strand, which is transcribed in the opposite

direction. Synthesis of mRNAs in adenovirus is actually much more complex than shown here. Many of the mRNAs shown for the upper strand are initially synthesized as a single, long transcript (25,000 nucleotides), which is then extensively processed to produce the separate mRNAs. Adenovirus causes upper respiratory tract infections in some vertebrates.

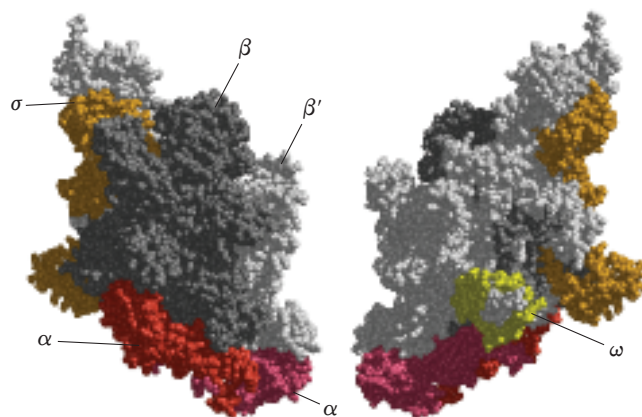


FIGURE 26-4 Structure of the RNA polymerase holoenzyme of the bacterium *Thermus aquaticus*. (Derived from PDB ID 1IW7.) The overall structure of this enzyme is very similar to that of the *E. coli* RNA polymerase; no DNA or RNA is shown here. The β subunit is in gray, the β' subunit is white; the two α subunits are different shades of red; the ω subunit is yellow; the σ subunit is orange. The image on the left is oriented as in Figure 26-6. When the structure is rotated 180° about the y axis (right) the small ω subunit is visible.

oriC consensus sequence; see Fig. 25-11). The consensus sequence at the -10 region is (5')TATAAT(3'); the consensus sequence at the -35 region is (5')TTGACA(3'). A third AT-rich recognition element, called the UP (upstream promoter) element, occurs between positions -40 and -60 in the promoters of cer-

tain highly expressed genes. The UP element is bound by the α subunit of RNA polymerase. The efficiency with which an RNA polymerase binds to a promoter and initiates transcription is determined in large measure by these sequences, the spacing between them, and their distance from the transcription start site.

Many independent lines of evidence attest to the functional importance of the sequences in the -35 and -10 regions. Mutations that affect the function of a given promoter often involve a base pair in these regions. Variations in the consensus sequence also affect the efficiency of RNA polymerase binding and transcription initiation. A change in only one base pair can decrease the rate of binding by several orders of magnitude. The promoter sequence thus establishes a basal level of expression that can vary greatly from one *E. coli* gene to the next. A method that provides information about the interaction between RNA polymerase and promoters is illustrated in Box 26-1.

The pathway of transcription initiation is becoming much better defined (Fig. 26-6a). It consists of two major parts, binding and initiation, each with multiple steps. First, the polymerase binds to the promoter, forming, in succession, a closed complex (in which the bound DNA is intact) and an open complex (in which the bound DNA is intact and partially unwound near the -10 sequence). Second, transcription is initiated within the complex, leading to a conformational change that converts the complex to the elongation form, followed by movement of the transcription complex away from

	UP element	-35 Region	Spacer	-10 Region	Spacer	RNA start
Consensus sequence	NNAAA ^{AA} _{TT} -A ^A _T TTTNNAAAANN	TTGACA	N ₁₇	TATAAT	N ₆	+1
<i>rrnB</i> P1	AGAAAATTATTTTAAATTCCT	GTGTCA	N ₁₆	TATAAT	N ₈	A
<i>trp</i>		TTGACA	N ₁₇	TTAACT	N ₇	A
<i>lac</i>		TTTACA	N ₁₇	TATGTT	N ₆	A
<i>recA</i>		TTGATA	N ₁₆	TATAAT	N ₇	A
<i>araBAD</i>		CTGACG	N ₁₈	TACTGT	N ₆	A

FIGURE 26-5 Typical *E. coli* promoters recognized by an RNA polymerase holoenzyme containing σ^{70} . Sequences of the nontemplate strand are shown, read in the 5'→3' direction, as is the convention for representations of this kind. The sequences vary from one promoter to the next, but comparisons of many promoters reveal similarities, particularly in the -10 and -35 regions. The sequence element UP, not present in all *E. coli* promoters, is shown in the P1 promoter for the highly expressed rRNA gene *rrnB*. UP elements, generally occur-

ring in the region between -40 and -60 , strongly stimulate transcription at the promoters that contain them. The UP element in the *rrnB* P1 promoter encompasses the region between -38 and -59 . The consensus sequence for *E. coli* promoters recognized by σ^{70} is shown second from the top. Spacer regions contain slightly variable numbers of nucleotides (N). Only the first nucleotide coding the RNA transcript (at position $+1$) is shown.

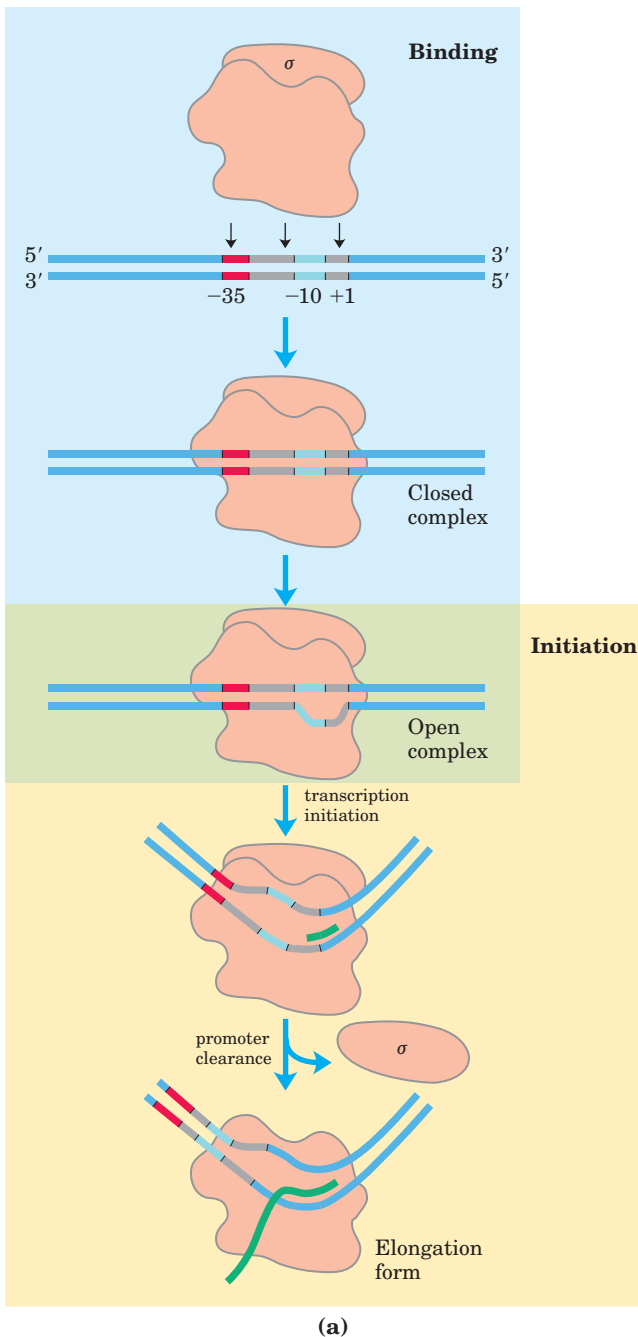
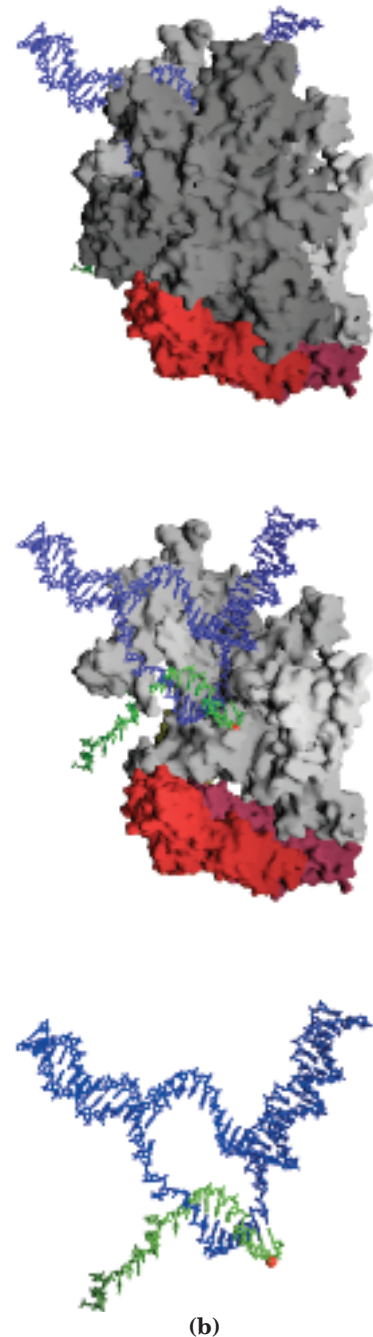


FIGURE 26-6 Transcription initiation and elongation by *E. coli* RNA polymerase. (a) Initiation of transcription requires several steps generally divided into two phases, binding and initiation. In the binding phase, the initial interaction of the RNA polymerase with the promoter leads to formation of a closed complex, in which the promoter DNA is stably bound but not unwound. A 12 to 15 bp region of DNA—from within the -10 region to position $+2$ or $+3$ —is then unwound to form an open complex. Additional intermediates (not shown) have been detected in the pathways leading to the closed and open complexes, along with several changes in protein conformation. The initiation phase encompasses transcription initiation and promoter clearance. Once the first 8 or 9 nucleotides of a new RNA are synthesized, the σ subunit is released and the polymerase leaves the promoter and becomes committed to elongation of the RNA.



(b) Structure of the RNA core polymerase from *E. coli*. RNA and DNA are included here to illustrate a polymerase in the elongation phase. Subunit coloring matches Figure 26-4: the β and β' subunits are light gray and white; the α subunits, shades of red. The ω subunit is on the opposite side of the complex and is not visible in this view. The σ subunit is not present in this complex, having dissociated after the initiation steps. The top panel shows the entire complex. The active site for transcription is in a cleft between the β and β' subunits. In the middle panel, the β subunit has been removed, exposing the active site and the DNA-RNA hybrid region. The active site is marked in part by a Mg^{2+} ion (red). In the bottom panel, all the protein has been removed to reveal the circuitous path taken by the DNA and RNA through the complex.

the promoter (promoter clearance). Any of these steps can be affected by the specific makeup of the promoter sequences. The σ subunit dissociates as the polymerase enters the elongation phase of transcription (Fig. 26-6a).

E. coli has other classes of promoters, bound by RNA polymerase holoenzymes with different σ subunits. An example is the promoters of the heat-shock genes. The products of this set of genes are made at higher levels when the cell has received an insult, such as a sudden increase in temperature. RNA polymerase binds to the promoters of these genes only when σ^{70} is replaced with the σ^{32} (M_r 32,000) subunit, which is specific for the heat-shock promoters (see Fig. 28-3). By using different σ subunits the cell can coordinate the expression of sets of genes, permitting major changes in cell physiology.

Transcription Is Regulated at Several Levels

Requirements for any gene product vary with cellular conditions or developmental stage, and transcription of each gene is carefully regulated to form gene products only in the proportions needed. Regulation can occur at any step in transcription, including elongation and termination. However, much of the regulation is directed at the polymerase binding and transcription initiation steps outlined in Figure 26-6. Differences in promoter sequences are just one of several levels of control.

The binding of proteins to sequences both near to and distant from the promoter can also affect levels of gene expression. Protein binding can *activate* transcription by facilitating either RNA polymerase binding or steps further along in the initiation process, or it can *repress* transcription by blocking the activity of the polymerase. In *E. coli*, one protein that activates transcription is the **cAMP receptor protein (CRP)**, which increases the transcription of genes coding for enzymes that metabolize sugars other than glucose when cells are grown in the absence of glucose. **Repressors** are proteins that block the synthesis of RNA at specific genes. In the case of the Lac repressor (Chapter 28), transcription of the genes for the enzymes of lactose metabolism is blocked when lactose is unavailable.

Transcription is the first step in the complicated and energy-intensive pathway of protein synthesis, so much of the regulation of protein levels in both bacterial and eukaryotic cells is directed at transcription, particularly its early stages. In Chapter 28 we describe many mechanisms by which this regulation is accomplished.

Specific Sequences Signal Termination of RNA Synthesis

RNA synthesis is processive (that is, the RNA polymerase has high processivity; p. 954)—necessarily so, because if an RNA polymerase released an RNA transcript prematurely, it could not resume synthesis of the same

RNA but instead would have to start over. However, an encounter with certain DNA sequences results in a pause in RNA synthesis, and at some of these sequences transcription is terminated. The process of termination is not yet well understood in eukaryotes, so our focus is again on bacteria. *E. coli* has at least two classes of termination signals: one class relies on a protein factor called ρ (rho) and the other is ρ -independent.

Most ρ -independent terminators have two distinguishing features. The first is a region that produces an RNA transcript with self-complementary sequences, permitting the formation of a hairpin structure (see Fig. 8-21a) centered 15 to 20 nucleotides before the projected end of the RNA strand. The second feature is a highly conserved string of three A residues in the template strand that are transcribed into U residues near the 3' end of the hairpin. When a polymerase arrives at a termination site with this structure, it pauses (Fig. 26-7). Formation of the hairpin structure in the RNA disrupts several A=U base pairs in the RNA-DNA hybrid segment and may disrupt important interactions

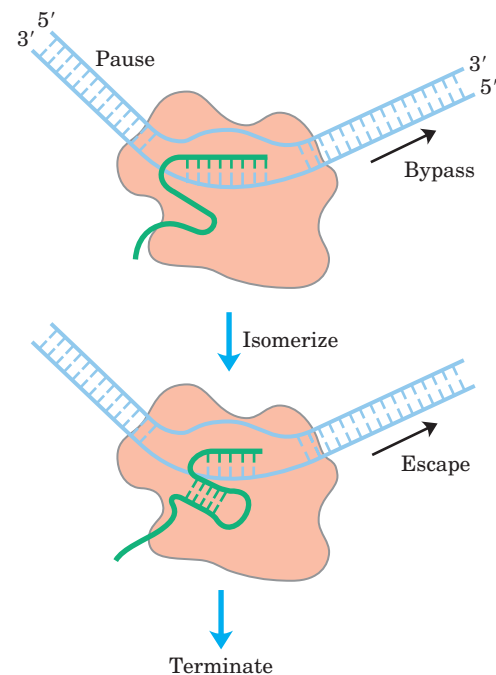


FIGURE 26-7 Model for ρ -independent termination of transcription in *E. coli*. RNA polymerase pauses at a variety of DNA sequences, some of which are terminators. One of two outcomes is then possible: the polymerase bypasses the site and continues on its way, or the complex undergoes a conformational change (isomerization). In the latter case, intramolecular pairing of complementary sequences in the newly formed RNA transcript may form a hairpin that disrupts the RNA-DNA hybrid and/or the interactions between the RNA and the polymerase, resulting in isomerization. An A=U hybrid region at the 3' end of the new transcript is relatively unstable, and the RNA dissociates completely, leading to termination and dissociation of the RNA molecule. This is the usual outcome at terminators. At other pause sites, the complex may escape after the isomerization step to continue RNA synthesis.

BOX 26-1 WORKING IN BIOCHEMISTRY

RNA Polymerase Leaves Its Footprint on a Promoter

Footprinting, a technique derived from principles used in DNA sequencing, identifies the DNA sequences bound by a particular protein. Researchers isolate a DNA fragment thought to contain sequences recognized by a DNA-binding protein and radiolabel

one end of one strand (Fig. 1). They then use chemical or enzymatic reagents to introduce random breaks in the DNA fragment (averaging about one per molecule). Separation of the labeled cleavage products (broken fragments of various lengths) by high-resolution electrophoresis produces a ladder of radioactive bands. In a separate tube, the cleavage procedure is

repeated on copies of the same DNA fragment in the presence of the DNA-binding protein. The researchers then subject the two sets of cleavage products to electrophoresis and compare them side by side. A gap (“footprint”) in the series of radioactive bands derived from the DNA-protein sample, attributable to protection of the DNA by the bound protein, identifies the sequences that the protein binds.

The precise location of the protein-binding site can be determined by directly sequencing (see Fig. 8-37) copies of the same DNA fragment and including the sequencing lanes (not shown here) on the same gel with the footprint. Figure 2 shows footprinting results for the binding of RNA polymerase to a DNA fragment containing a promoter. The polymerase covers 60 to 80 bp; protection by the bound enzyme includes the -10 and -35 regions.

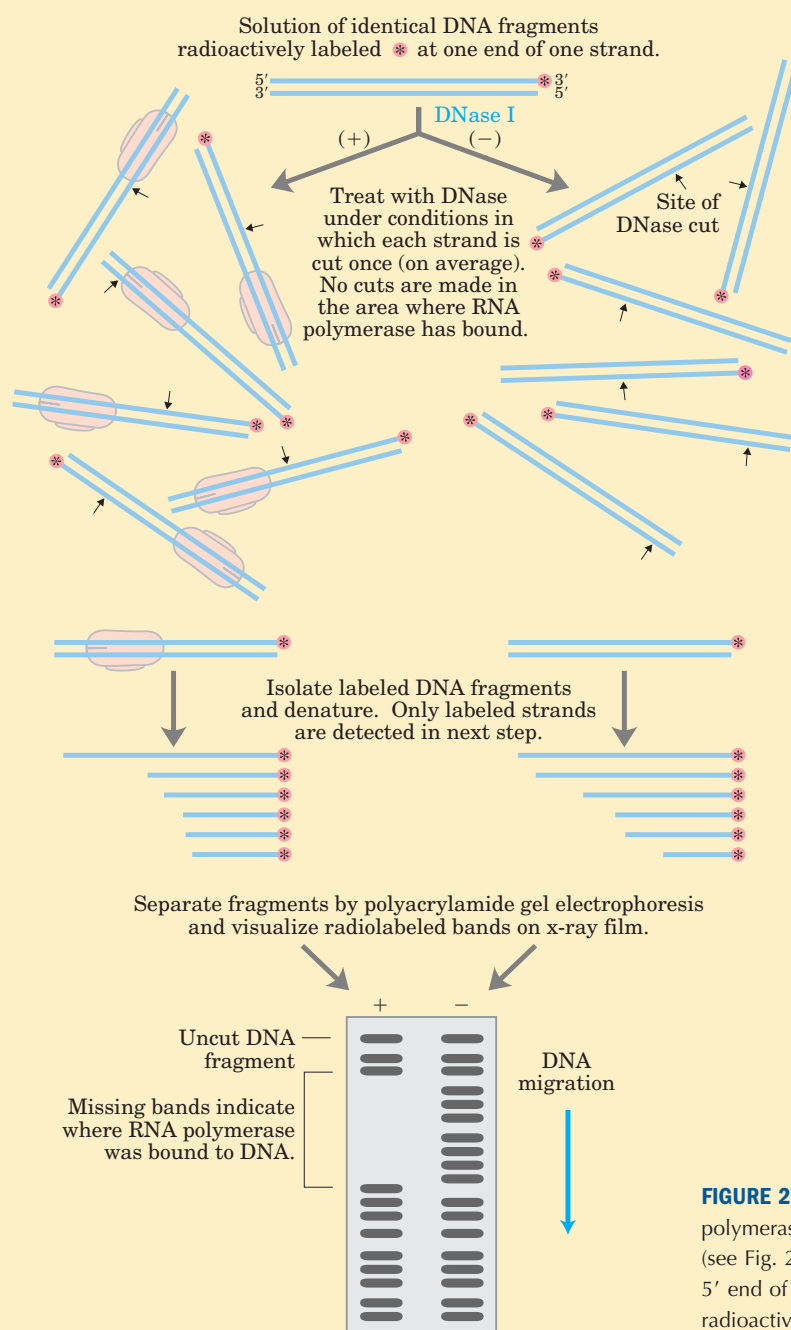


FIGURE 1 Footprint analysis of the RNA polymerase-binding site on a DNA fragment. Separate experiments are carried out in the presence (+) and absence (-) of the polymerase.

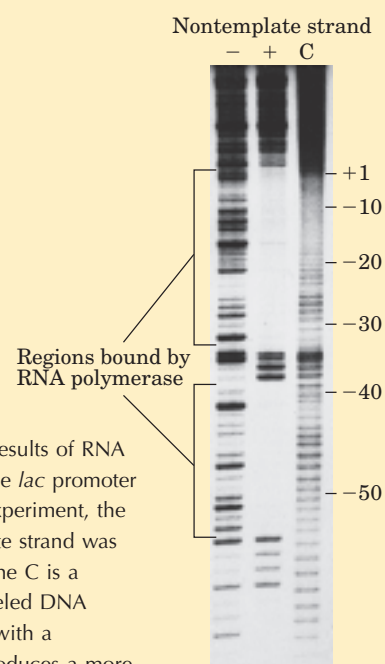


FIGURE 2 Footprinting results of RNA polymerase binding to the *lac* promoter (see Fig. 26-5). In this experiment, the 5' end of the nontemplate strand was radioactively labeled. Lane C is a control in which the labeled DNA fragments were cleaved with a chemical reagent that produces a more uniform banding pattern.

between RNA and the RNA polymerase, facilitating dissociation of the transcript.

The ρ -dependent terminators lack the sequence of repeated A residues in the template strand but usually include a CA-rich sequence called a *rut* (*rho* utilization) element. The ρ protein associates with the RNA at specific binding sites and migrates in the 5'→3' direction until it reaches the transcription complex that is paused at a termination site. Here it contributes to release of the RNA transcript. The ρ protein has an ATP-dependent RNA-DNA helicase activity that promotes translocation of the protein along the RNA, and ATP is hydrolyzed by ρ protein during the termination process. The detailed mechanism by which the protein promotes the release of the RNA transcript is not known.

Eukaryotic Cells Have Three Kinds of Nuclear RNA Polymerases

The transcriptional machinery in the nucleus of a eukaryotic cell is much more complex than that in bacteria. Eukaryotes have three RNA polymerases, designated I, II, and III, which are distinct complexes but have certain subunits in common. Each polymerase has a specific function and is recruited to a specific promoter sequence.

RNA polymerase I (Pol I) is responsible for the synthesis of only one type of RNA, a transcript called pre-ribosomal RNA (or pre-rRNA), which contains the precursor for the 18S, 5.8S, and 28S rRNAs (see Fig. 26–22). Pol I promoters vary greatly in sequence from one species to another. The principal function of RNA polymerase II (Pol II) is synthesis of mRNAs and some specialized RNAs. This enzyme can recognize thousands of promoters that vary greatly in sequence. Many Pol II promoters have a few sequence features in common, including a TATA box (eukaryotic consensus sequence TATAAA) near base pair –30 and an Inr sequence (initiator) near the RNA start site at +1 (Fig. 26–8).

RNA polymerase III (Pol III) makes tRNAs, the 5S rRNA, and some other small specialized RNAs. The pro-

motors recognized by Pol III are well characterized. Interestingly, some of the sequences required for the regulated initiation of transcription by Pol III are located within the gene itself, whereas others are in more conventional locations upstream of the RNA start site (Chapter 28).

RNA Polymerase II Requires Many Other Protein Factors for Its Activity

RNA polymerase II is central to eukaryotic gene expression and has been studied extensively. Although this polymerase is strikingly more complex than its bacterial counterpart, the complexity masks a remarkable conservation of structure, function, and mechanism. Pol II is a huge enzyme with 12 subunits. The largest subunit (RBP1) exhibits a high degree of homology to the β' subunit of bacterial RNA polymerase. Another subunit (RBP2) is structurally similar to the bacterial β subunit, and two others (RBP3 and RBP11) show some structural homology to the two bacterial α subunits. Pol II must function with genomes that are more complex and with DNA molecules more elaborately packaged than in bacteria. The need for protein-protein contacts with the numerous other protein factors required to navigate this labyrinth accounts in large measure for the added complexity of the eukaryotic polymerase.

The largest subunit of Pol II also has an unusual feature, a long carboxyl-terminal tail consisting of many repeats of a consensus heptad amino acid sequence –YSPTSPS–. There are 27 repeats in the yeast enzyme (18 exactly matching the consensus) and 52 (21 exact) in the mouse and human enzymes. This carboxyl-terminal domain (CTD) is separated from the main body of the enzyme by an unstructured linker sequence. The CTD has many important roles in Pol II function, as outlined below.

RNA polymerase II requires an array of other proteins, called **transcription factors**, in order to form the active transcription complex. The **general transcription factors** required at every Pol II promoter

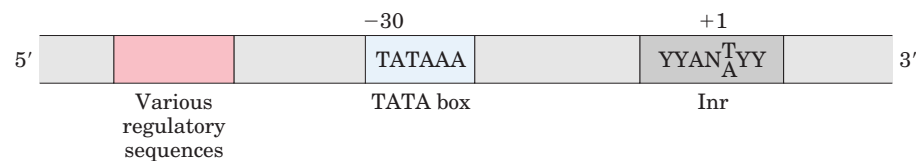


FIGURE 26–8 Common sequences in promoters recognized by eukaryotic RNA polymerase II. The TATA box is the major assembly point for the proteins of the preinitiation complexes of Pol II. The DNA is unwound at the initiator sequence (Inr), and the transcription start site is usually within or very near this sequence. In the Inr consensus sequence shown here, N represents any nucleotide; Y, a pyrimidine nucleotide. Many additional sequences serve as binding sites for a wide variety of proteins that affect the activity of Pol II. These sequences are important in regulating Pol II promoters and vary greatly in type and

number, and in general the eukaryotic promoter is much more complex than suggested here. Many of the sequences are located within a few hundred base pairs of the TATA box on the 5' side; others may be thousands of base pairs away. The sequence elements summarized here are more variable among the Pol II promoters of eukaryotes than among the *E. coli* promoters (see Fig. 26–5). Many Pol II promoters lack a TATA box or a consensus Inr element or both. Additional sequences around the TATA box and downstream (to the right as drawn) of Inr may be recognized by one or more transcription factors.

(factors usually designated TFII with an additional identifier) are highly conserved in all eukaryotes (Table 26–1). The process of transcription by Pol II can be described in terms of several phases—assembly, initiation, elongation, termination—each associated with characteristic proteins (Fig. 26–9). The step-by-step pathway

described below leads to active transcription in vitro. In the cell, many of the proteins may be present in larger, preassembled complexes, simplifying the pathways for assembly on promoters. As you read about this process, consult Figure 26–9 and Table 26–1 to help keep track of the many participants.

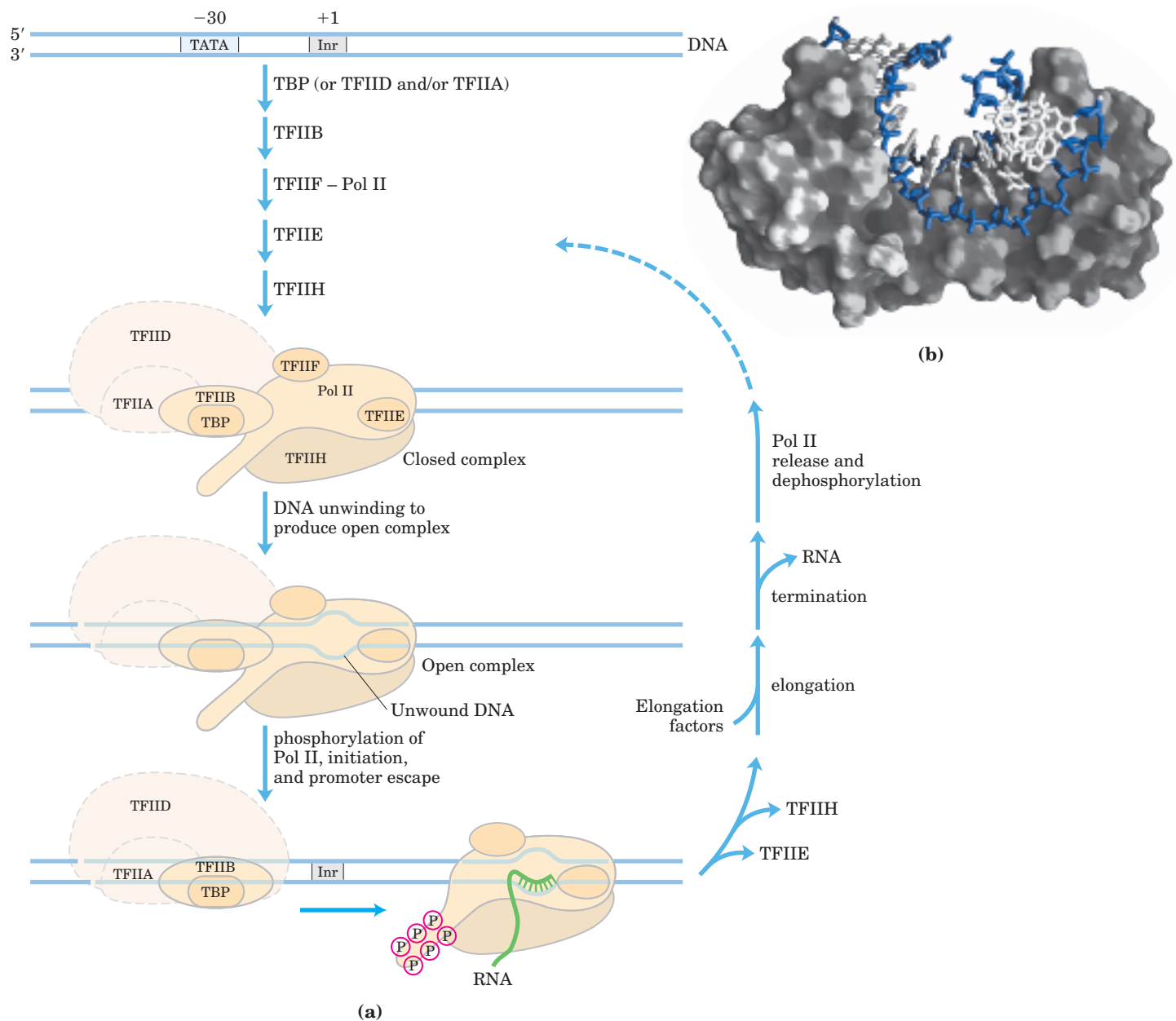


FIGURE 26–9 Transcription at RNA polymerase II promoters. (a) The sequential assembly of TBP (often with TFIIA), TFIIB, TFIIF plus Pol II, TFIIE, and TFIIH results in a closed complex. TBP often binds as part of a larger complex, TFIID. Some of the TFIID subunits play a role in transcription regulation (see Fig. 28–30). Within the complex, the DNA is unwound at the Inr region by the helicase activity of TFIIH and perhaps of TFIIE, creating an open complex. The carboxyl-terminal do-

main of the largest Pol II subunit is phosphorylated by TFIIH, and the polymerase then escapes the promoter and begins transcription. Elongation is accompanied by the release of many transcription factors and is also enhanced by elongation factors (see Table 26–1). After termination, Pol II is released, dephosphorylated, and recycled. (b) The structure of human TBP (gray) bound to DNA (blue and white) (PDB ID 1TGH).

TABLE 26-1 Proteins Required for Initiation of Transcription at the RNA Polymerase II (Pol II) Promoters of Eukaryotes

<i>Transcription protein</i>	<i>Number of subunits</i>	<i>Subunit(s) M_r</i>	<i>Function(s)</i>
Initiation			
Pol II	12	10,000–220,000	Catalyzes RNA synthesis
TBP (TATA-binding protein)	1	38,000	Specifically recognizes the TATA box
TFIIA	3	12,000, 19,000, 35,000	Stabilizes binding of TFIIB and TBP to the promoter
TFIIB	1	35,000	Binds to TBP; recruits Pol II–TFIIF complex
TFIIE	2	34,000, 57,000	Recruits TFIIH; has ATPase and helicase activities
TFIIF	2	30,000, 74,000	Binds tightly to Pol II; binds to TFIIB and prevents binding of Pol II to nonspecific DNA sequences
TFIIH	12	35,000–89,000	Unwinds DNA at promoter (helicase activity); phosphorylates Pol II (within the CTD); recruits nucleotide-excision repair proteins
Elongation*			
ELL [†]	1	80,000	Phosphorylates Pol II (within the CTD)
p-TEFb	2	43,000, 124,000	
SII (TFIIS)	1	38,000	
Elongin (SIII)	3	15,000, 18,000, 110,000	

*The function of all elongation factors is to suppress the pausing or arrest of transcription by the Pol II–TFIIF complex.

[†]Name derived from eleven-nineteen lysine-rich leukemia. The gene for ELL is the site of chromosomal recombination events frequently associated with acute myeloid leukemia.

Assembly of RNA Polymerase and Transcription Factors at a Promoter

The formation of a closed complex begins when the TATA-binding protein (TBP) binds to the TATA box (Fig. 26-9b). TBP is bound in turn by the transcription factor TFIIB, which also binds to DNA on either side of TBP. TFIIA binding, although not always essential, can stabilize the TFIIB-TBP complex on the DNA and can be important at nonconsensus promoters where TBP binding is relatively weak. The TFIIB-TBP complex is next bound by another complex consisting of TFIIF and Pol II. TFIIF helps target Pol II to its promoters, both by interacting with TFIIB and by reducing the binding of the polymerase to nonspecific sites on the DNA. Finally, TFIIE and TFIIH bind to create the closed complex. TFIIH has DNA helicase activity that promotes the unwinding of DNA near the RNA start site (a process requiring the hydrolysis of ATP), thereby creating an open complex. Counting all the subunits of the various essential factors (excluding TFIIA), this minimal active assembly has more than 30 polypeptides.

RNA Strand Initiation and Promoter Clearance TFIIH has an additional function during the initiation phase. A kinase activity in one of its subunits phosphorylates Pol II at many places in the CTD (Fig. 26-9). Several other protein kinases, including CDK9 (cyclin-dependent kinase 9), which is part of the complex pTEFb (positive transcription elongation factor *b*), also phosphorylate the

CTD. This causes a conformational change in the overall complex, initiating transcription. Phosphorylation of the CTD is also important during the subsequent elongation phase, and it affects the interactions between the transcription complex and other enzymes involved in processing the transcript (as described below).

During synthesis of the initial 60 to 70 nucleotides of RNA, first TFIIE and then TFIIH is released, and Pol II enters the elongation phase of transcription.

Elongation, Termination, and Release TFIIF remains associated with Pol II throughout elongation. During this stage, the activity of the polymerase is greatly enhanced by proteins called elongation factors (Table 26-1). The elongation factors suppress pausing during transcription and also coordinate interactions between protein complexes involved in the posttranscriptional processing of mRNAs. Once the RNA transcript is completed, transcription is terminated. Pol II is dephosphorylated and recycled, ready to initiate another transcript (Fig. 26-9).

Regulation of RNA Polymerase II Activity Regulation of transcription at Pol II promoters is quite elaborate. It involves the interaction of a wide variety of other proteins with the preinitiation complex. Some of these regulatory proteins interact with transcription factors, others with Pol II itself. Many interact through TFIID, a complex of about 12 proteins, including TBP and certain

TBP-associated factors, or TAFs. The regulation of transcription is described in more detail in Chapter 28.

Diverse Functions of TFIIF In eukaryotes, the repair of damaged DNA (see Table 25–5) is more efficient within genes that are actively being transcribed than for other damaged DNA, and the template strand is repaired somewhat more efficiently than the nontemplate strand. These remarkable observations are explained by the alternative roles of the TFIIF subunits. Not only does TFIIF participate in the formation of the closed complex during assembly of a transcription complex (as described above), but some of its subunits are also essential components of the separate nucleotide-excision repair complex (see Fig. 25–24).

When Pol II transcription halts at the site of a DNA lesion, TFIIF can interact with the lesion and recruit the entire nucleotide-excision repair complex. Genetic loss of certain TFIIF subunits can produce human diseases. Some examples are xeroderma pigmentosum (see Box 25–1) and Cockayne's syndrome, which is characterized by arrested growth, photosensitivity, and neurological disorders. ■

DNA-Dependent RNA Polymerase Undergoes Selective Inhibition

The elongation of RNA strands by RNA polymerase in both bacteria and eukaryotes is inhibited by the antibiotic **actinomycin D** (Fig. 26–10). The planar portion of this molecule inserts (intercalates) into the double-helical DNA between successive G≡C base pairs, deforming the DNA. This prevents movement of the polymerase along the template. Because actinomycin D

inhibits RNA elongation in intact cells as well as in cell extracts, it is used to identify cell processes that depend on RNA synthesis. **Acridine** inhibits RNA synthesis in a similar fashion (Fig. 26–10).

Rifampicin inhibits bacterial RNA synthesis by binding to the β subunit of bacterial RNA polymerases, preventing the promoter clearance step of transcription (Fig. 26–6). It is sometimes used as an antibiotic.

The mushroom *Amanita phalloides* has evolved a very effective defense mechanism against predators. It produces **α -amanitin**, which disrupts mRNA formation in animal cells by blocking Pol II and, at higher concentrations, Pol III. Neither Pol I nor bacterial RNA polymerase is sensitive to α -amanitin—nor is the RNA polymerase II of *A. phalloides* itself!

SUMMARY 26.1 DNA-Dependent Synthesis of RNA

- Transcription is catalyzed by DNA-dependent RNA polymerases, which use ribonucleoside 5'-triphosphates to synthesize RNA complementary to the template strand of duplex DNA. Transcription occurs in several phases: binding of RNA polymerase to a DNA site called a promoter, initiation of transcript synthesis, elongation, and termination.
- Bacterial RNA polymerase requires a special subunit to recognize the promoter. As the first committed step in transcription, binding of RNA polymerase to the promoter and initiation of transcription are closely regulated. Transcription stops at sequences called terminators.

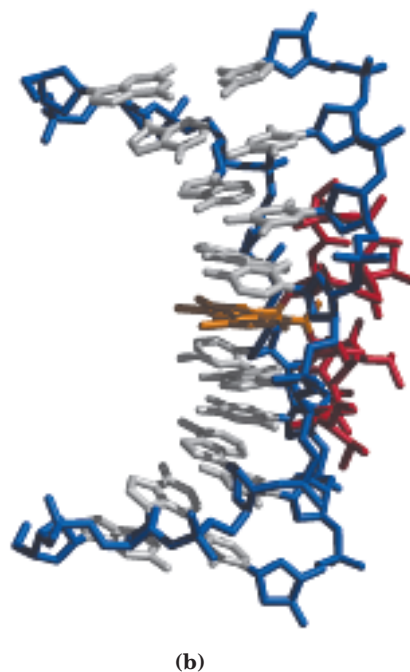
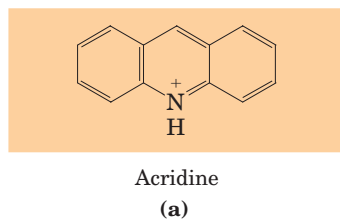
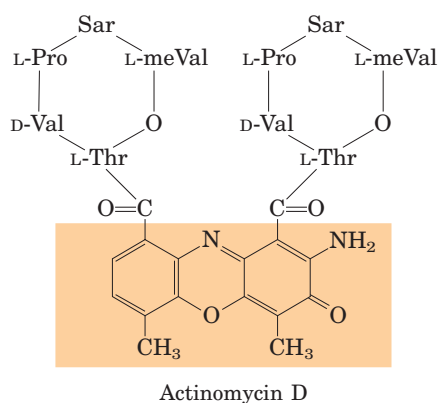


FIGURE 26–10 Actinomycin D and acridine, inhibitors of DNA transcription. (a) The shaded portion of actinomycin D is planar and intercalates between two successive G≡C base pairs in duplex DNA. The two cyclic peptide structures of actinomycin D bind to the minor groove of the double helix. Sarcosine (Sar) is *N*-methylglycine; meVal is methylvaline. Acridine also acts by intercalation in DNA. (b) A complex of actinomycin D with DNA (PDB ID 1DSC). The DNA backbone is shown in blue, the bases are white, the intercalated part of actinomycin (shaded in (a)) is orange, and the remainder of the actinomycin is red. The DNA is bent as a result of the actinomycin binding.

■ Eukaryotic cells have three types of RNA polymerases. Binding of RNA polymerase II to its promoters requires an array of proteins called transcription factors. Elongation factors participate in the elongation phase of transcription. The largest subunit of Pol II has a long carboxyl-terminal domain, which is phosphorylated during the initiation and elongation phases.

26.2 RNA Processing

Many of the RNA molecules in bacteria and virtually all RNA molecules in eukaryotes are processed to some degree after synthesis. Some of the most interesting molecular events in RNA metabolism occur during this postsynthetic processing. Intriguingly, several of the enzymes that catalyze these reactions consist of RNA rather than protein. The discovery of these catalytic RNAs, or **ribozymes**, has brought a revolution in thinking about RNA function and about the origin of life.

A newly synthesized RNA molecule is called a **primary transcript**. Perhaps the most extensive processing of primary transcripts occurs in eukaryotic mRNAs and in tRNAs of both bacteria and eukaryotes.

The primary transcript for a eukaryotic mRNA typically contains sequences encompassing one gene, although the sequences encoding the polypeptide may not be contiguous. Noncoding tracts that break up the coding region of the transcript are called introns, and the coding segments are called exons (see the discussion of introns and exons in DNA in Chapter 24). In a process called **splicing**, the introns are removed from the primary transcript and the exons are joined to form a con-

tinuous sequence that specifies a functional polypeptide. Eukaryotic mRNAs are also modified at each end. A modified residue called a 5' cap (p. 1008) is added at the 5' end. The 3' end is cleaved, and 80 to 250 A residues are added to create a poly(A) "tail." The sometimes elaborate protein complexes that carry out each of these three mRNA-processing reactions do not operate independently. They appear to be organized in association with each other and with the phosphorylated CTD of Pol II; each complex affects the function of the others. Other proteins involved in mRNA transport to the cytoplasm are also associated with the mRNA in the nucleus, and the processing of the transcript is coupled to its transport. In effect, a eukaryotic mRNA, as it is synthesized, is ensconced in an elaborate complex involving dozens of proteins. The composition of the complex changes as the primary transcript is processed, transported to the cytoplasm, and delivered to the ribosome for translation. These processes are outlined in Figure 26–11 and described in more detail below.

The primary transcripts of prokaryotic and eukaryotic tRNAs are processed by the removal of sequences from each end (cleavage) and in a few cases by the removal of introns (splicing). Many bases and sugars in tRNAs are also modified; mature tRNAs are replete with unusual bases not found in other nucleic acids (see Fig. 26–24).

The ultimate fate of any RNA is its complete and regulated degradation. The rate of turnover of RNAs plays a critical role in determining their steady-state levels and the rate at which cells can shut down expression of a gene whose product is no longer needed. During the development of multicellular organisms, for example, certain proteins must be expressed at one stage only, and the mRNA encoding such a protein must be made and destroyed at the appropriate times.

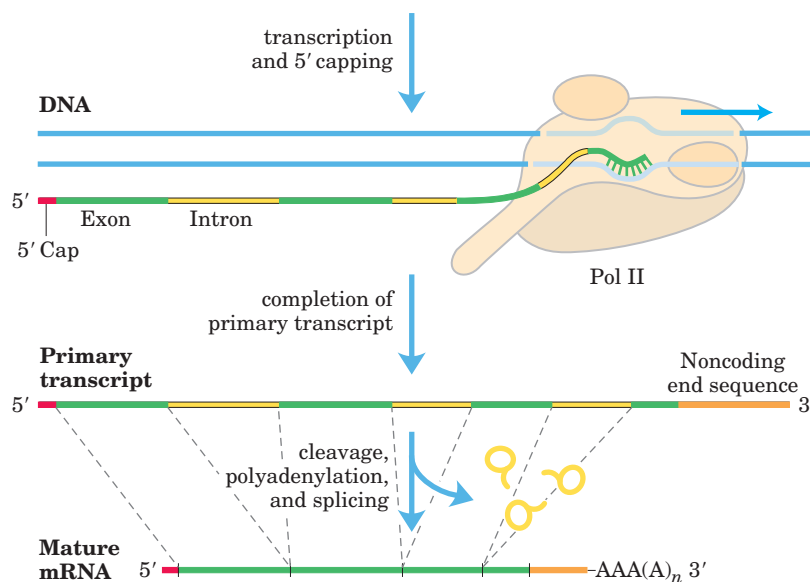


FIGURE 26–11 Formation of the primary transcript and its processing during maturation of mRNA in a eukaryotic cell. The 5' cap (red) is added before synthesis of the primary transcript is complete. A noncoding sequence following the last exon is shown in orange. Splicing can occur either before or after the cleavage and polyadenylation steps. All the processes shown here take place within the nucleus.

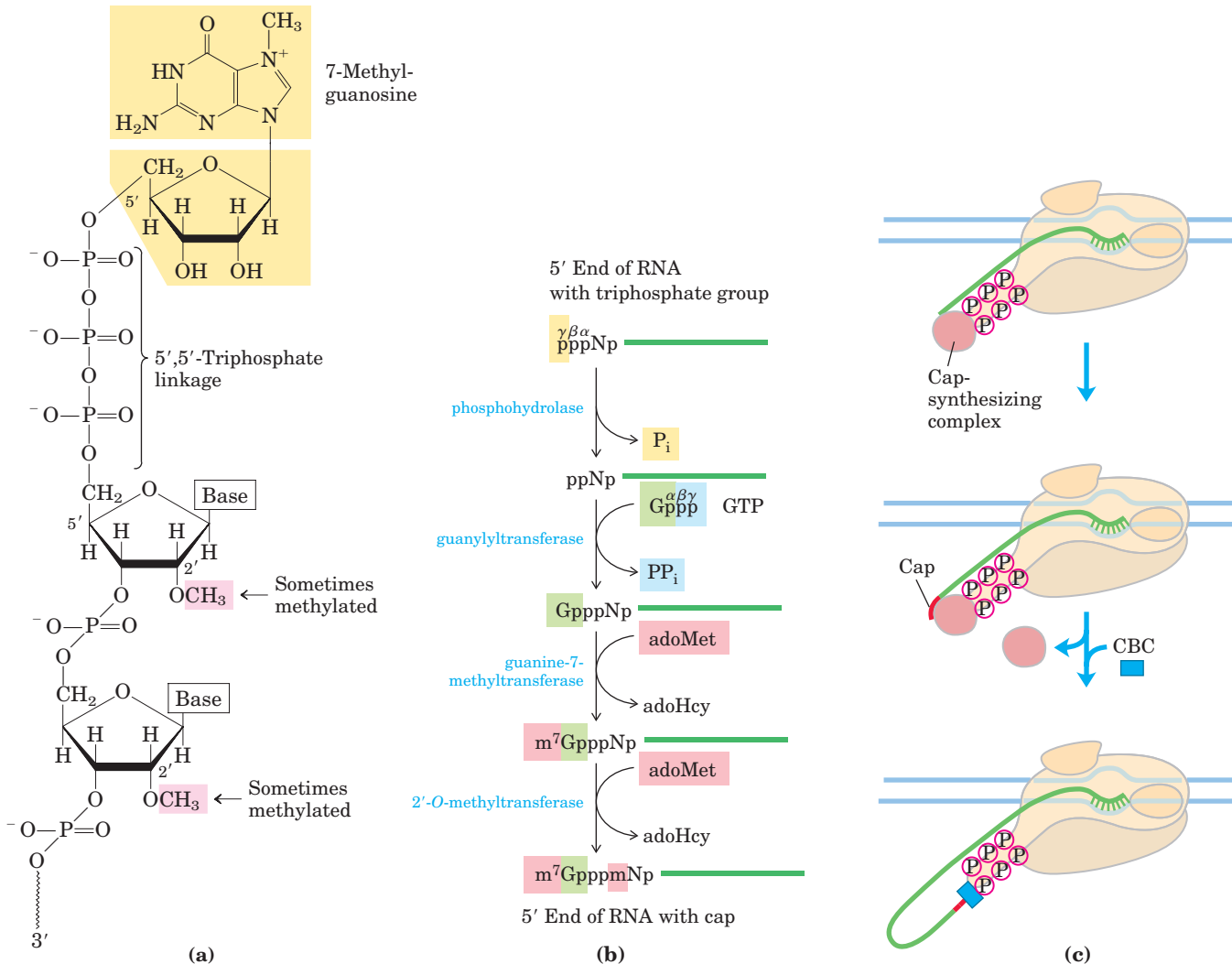


FIGURE 26-12 The 5' cap of mRNA. (a) 7-Methylguanosine is joined to the 5' end of almost all eukaryotic mRNAs in an unusual 5',5'-triphosphate linkage. Methyl groups (pink) are often found at the 2' position of the first and second nucleotides. RNAs in yeast cells lack the 2'-methyl groups. The 2'-methyl group on the second nucleotide

is generally found only in RNAs from vertebrate cells. (b) Generation of the 5' cap involves four to five separate steps (adoHcy is *S*-adenosylhomocysteine). (c) Synthesis of the cap is carried out by enzymes tethered to the CTD of Pol II. The cap remains tethered to the CTD through an association with the cap-binding complex (CBC).

Eukaryotic mRNAs Are Capped at the 5' End

Most eukaryotic mRNAs have a **5' cap**, a residue of 7-methylguanosine linked to the 5'-terminal residue of the mRNA through an unusual 5',5'-triphosphate linkage (Fig. 26-12). The 5' cap helps protect mRNA from ribonucleases. The cap also binds to a specific cap-binding complex of proteins and participates in binding of the mRNA to the ribosome to initiate translation (Chapter 27).

The 5' cap is formed by condensation of a molecule of GTP with the triphosphate at the 5' end of the transcript. The guanine is subsequently methylated at N-7, and additional methyl groups are often added at the 2' hydroxyls of the first and second nucleotides adjacent to the cap (Fig. 26-12). The methyl groups are derived from *S*-adenosylmethionine. All these reactions occur

very early in transcription, after the first 20 to 30 nucleotides of the transcript have been added. All three of the capping enzymes, and through them the 5' end of the transcript itself, are associated with the RNA polymerase II CTD until the cap is synthesized. The capped 5' end is then released from the capping enzymes and bound by the cap-binding complex (Fig. 26-12c).

Both Introns and Exons Are Transcribed from DNA into RNA

In bacteria, a polypeptide chain is generally encoded by a DNA sequence that is colinear with the amino acid sequence, continuing along the DNA template without interruption until the information needed to specify the polypeptide is complete. However, the notion that *all* genes are continuous was disproved in 1977 when

Phillip Sharp and Richard Roberts independently discovered that many genes for polypeptides in eukaryotes are interrupted by noncoding sequences (introns).

The vast majority of genes in vertebrates contain introns; among the few exceptions are those that encode histones. The occurrence of introns in other eukaryotes varies. Many genes in the yeast *Saccharomyces cerevisiae* lack introns, although in some other yeast species introns are more common. Introns are also found in a few eubacterial and archaeobacterial genes. Introns in DNA are transcribed along with the rest of the gene by RNA polymerases. The introns in the primary RNA transcript are then spliced, and the exons are joined to form a mature, functional RNA. In eukaryotic mRNAs, most exons are less than 1,000 nucleotides long, with many in the 100 to 200 nucleotide size range, encoding stretches of 30 to 60 amino acids within a longer polypeptide. Introns vary in size from 50 to 20,000 nucleotides. Genes of higher eukaryotes, including humans, typically have much more DNA devoted to introns than to exons. Many genes have introns; some genes have dozens of them.

RNA Catalyzes the Splicing of Introns

There are four classes of introns. The first two, the group I and group II introns, differ in the details of their splicing mechanisms but share one surprising characteristic: they are *self-splicing*—no protein enzymes are involved. Group I introns are found in some nuclear, mitochondrial, and chloroplast genes coding for rRNAs, mRNAs, and tRNAs. Group II introns are generally found in the primary transcripts of mitochondrial or chloroplast mRNAs in fungi, algae, and plants. Group I and group II introns are also found among the rarer examples of introns in bacteria. Neither class requires a high-energy cofactor (such as ATP) for splicing. The splicing

mechanisms in both groups involve two transesterification reaction steps (Fig. 26–13). A ribose 2'- or 3'-hydroxyl group makes a nucleophilic attack on a phosphorus and, in each step, a new phosphodiester bond is formed at the expense of the old, maintaining the balance of energy. These reactions are very similar to the DNA breaking and rejoining reactions promoted by topoisomerases (see Fig. 24–21) and site-specific recombinases (see Fig. 25–38).

The group I splicing reaction requires a guanine nucleoside or nucleotide cofactor, but the cofactor is not used as a source of energy; instead, the 3'-hydroxyl group of guanosine is used as a nucleophile in the first step of the splicing pathway. The guanosine 3'-hydroxyl group forms a normal 3',5'-phosphodiester bond with the 5' end of the intron (Fig. 26–14). The 3' hydroxyl of the exon that is displaced in this step then acts as a nucleophile in a similar reaction at the 3' end of the intron. The result is precise excision of the intron and ligation of the exons.

In group II introns the reaction pattern is similar except for the nucleophile in the first step, which in this case is the 2'-hydroxyl group of an A residue *within* the intron (Fig. 26–15). A branched lariat structure is formed as an intermediate.

Self-splicing of introns was first revealed in 1982 in studies of the splicing mechanism of the group I rRNA intron from the ciliated protozoan *Tetrahymena thermophila*, conducted by Thomas Cech and colleagues. These workers transcribed isolated *Tetrahymena* DNA (including the intron) in vitro using purified bacterial RNA polymerase. The resulting RNA spliced itself accurately without any protein enzymes from *Tetrahymena*. The discovery that RNAs could have catalytic functions was a milestone in our understanding of biological systems.

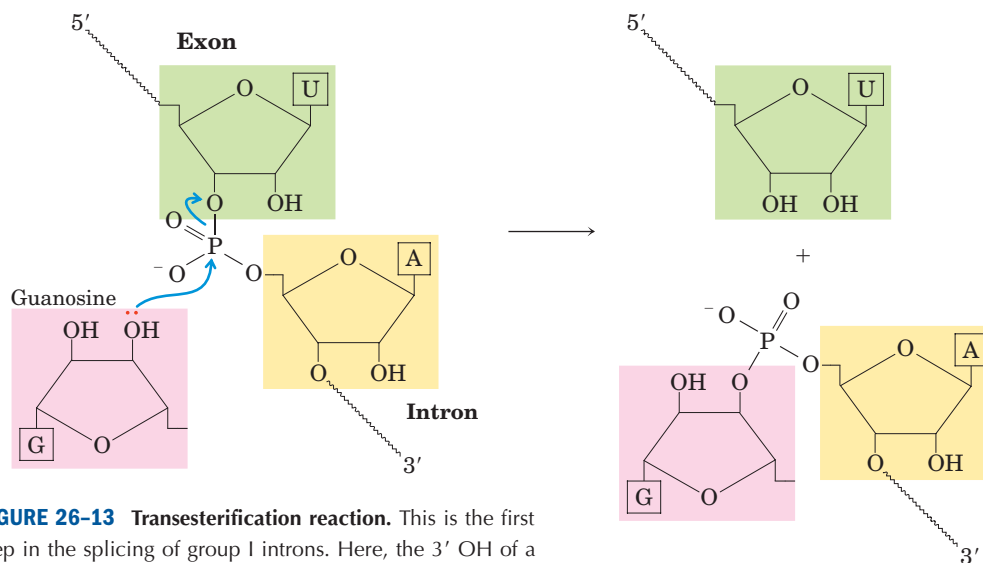


FIGURE 26–13 Transesterification reaction. This is the first step in the splicing of group I introns. Here, the 3' OH of a guanosine molecule acts as nucleophile.



Thomas Cech

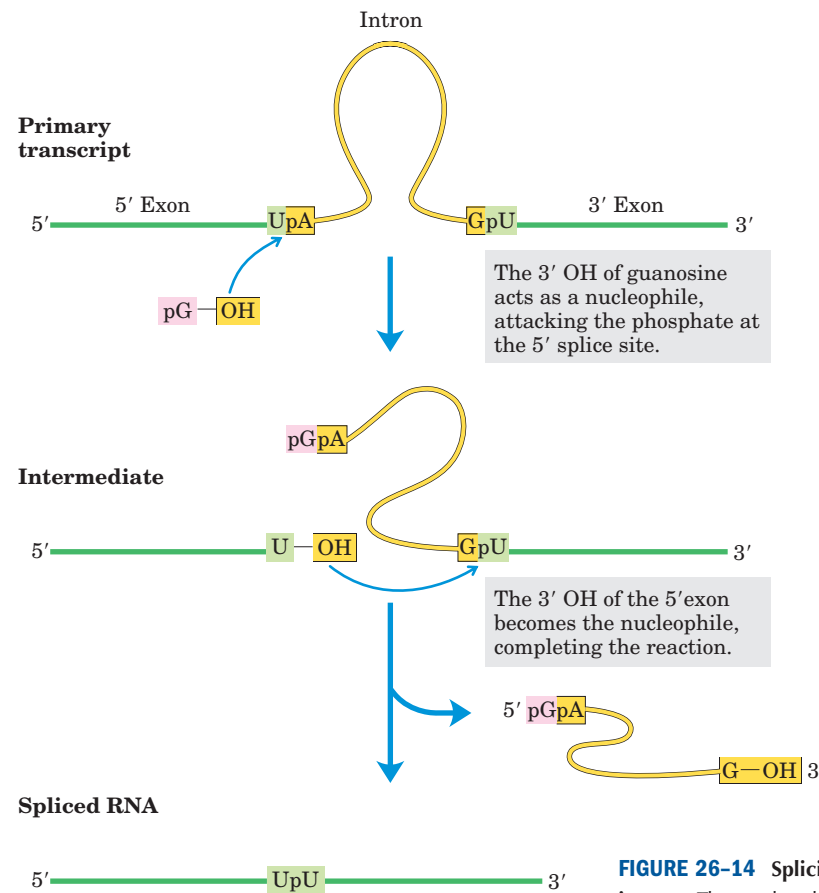



FIGURE 26-14 Splicing mechanism of group I introns. The nucleophile in the first step may be guanosine, GMP, GDP, or GTP. The spliced intron is eventually degraded.

Most introns are *not* self-splicing, and these types are not designated with a group number. The third and largest class of introns includes those found in nuclear mRNA primary transcripts. These are called **spliceosomal introns**, because their removal occurs within and is catalyzed by a large protein complex called a **spliceosome**. Within the spliceosome, the introns undergo splicing by the same lariat-forming mechanism as the group II introns. The spliceosome is made up of specialized RNA-protein complexes, small nuclear ribonucleoproteins (snRNPs, often pronounced “snurps”). Each snRNP contains one of a class of eukaryotic RNAs, 100 to 200 nucleotides long, known as **small nuclear RNAs (snRNAs)**. Five snRNAs (U1, U2, U4, U5, and U6) involved in splicing reactions are generally found in abundance in eukaryotic nuclei. The RNAs and proteins in snRNPs are highly conserved in eukaryotes from yeasts to humans.  **mRNA Splicing**

Spliceosomal introns generally have the dinucleotide sequence GU and AG at the 5' and 3' ends, respectively, and these sequences mark the sites where splicing occurs. The U1 snRNA contains a sequence complementary to sequences near the 5' splice site of nuclear mRNA introns (Fig. 26-16a), and the U1 snRNP

binds to this region in the primary transcript. Addition of the U2, U4, U5, and U6 snRNPs leads to formation of the spliceosome (Fig. 26-16b). The snRNPs together contribute five RNAs and about 50 proteins to the spliceosome, a supramolecular assembly nearly as complex as the ribosome (described in Chapter 27). ATP is required for assembly of the spliceosome, but the RNA cleavage-ligation reactions do not seem to require ATP. Some mRNA introns are spliced by a less common type of spliceosome, in which the U1 and U2 snRNPs are replaced by the U11 and U12 snRNPs. Whereas U1- and U2-containing spliceosomes remove introns with (5')GU and AG(3') terminal sequences, as shown in Figure 26-16, the U11- and U12-containing spliceosomes remove a rare class of introns that have (5')AU and AC(3') terminal sequences to mark the intronic splice sites. The spliceosomes used in nuclear RNA splicing may have evolved from more ancient group II introns, with the snRNPs replacing the catalytic domains of their self-splicing ancestors.

Some components of the splicing apparatus appear to be tethered to the CTD of RNA polymerase II, suggesting an interesting model for the splicing reaction. As the first splice junction is synthesized, it is bound by

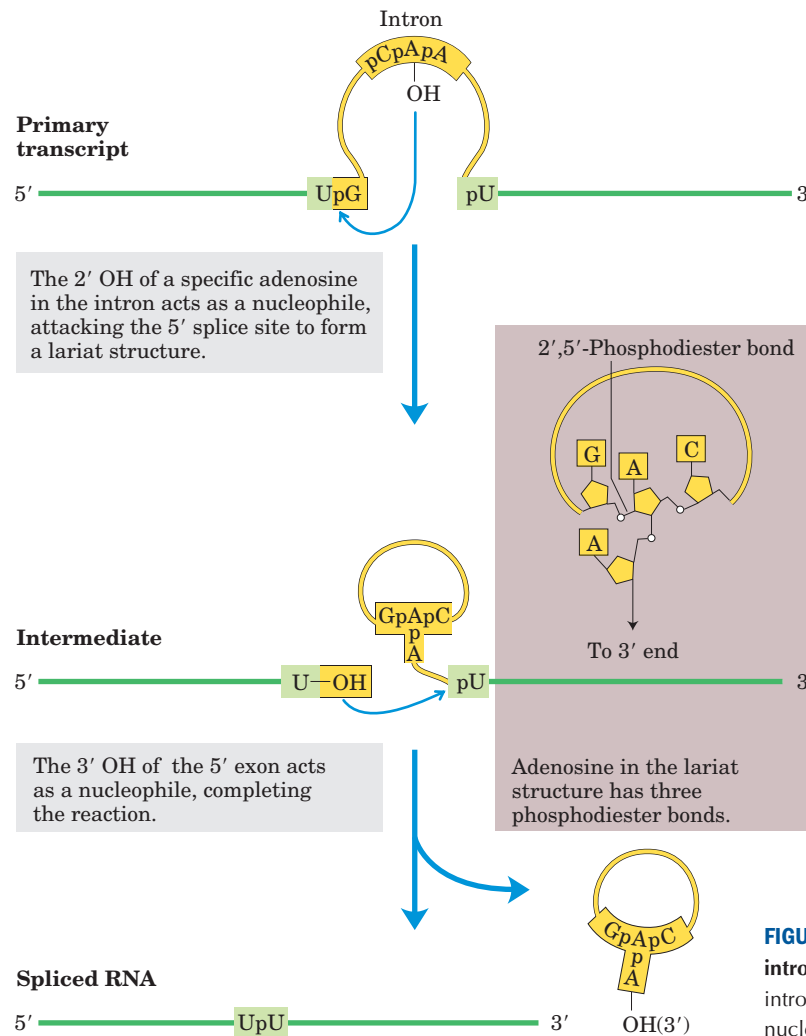


FIGURE 26-15 Splicing mechanism of group II introns. The chemistry is similar to that of group I intron splicing, except for the identity of the nucleophile in the first step and formation of a lariatlike intermediate, in which one branch is a 2',5'-phosphodiester bond.

a tethered spliceosome. The second splice junction is then captured by this complex as it passes, facilitating the juxtaposition of the intron ends and the subsequent splicing process (Fig. 26-16c). After splicing, the intron remains in the nucleus and is eventually degraded.

The fourth class of introns, found in certain tRNAs, is distinguished from the group I and II introns in that the splicing reaction requires ATP and an endonuclease. The splicing endonuclease cleaves the phosphodiester bonds at both ends of the intron, and the two exons are joined by a mechanism similar to the DNA ligase reaction (see Fig. 25-16).

Although spliceosomal introns appear to be limited to eukaryotes, the other intron classes are not. Genes with group I and II introns have now been found in both bacteria and bacterial viruses. Bacteriophage T4, for example, has several protein-encoding genes with group I introns. Introns appear to be more common in archaeobacteria than in eubacteria.

Eukaryotic mRNAs Have a Distinctive 3' End Structure

At their 3' end, most eukaryotic mRNAs have a string of 80 to 250 A residues, making up the **poly(A) tail**. This tail serves as a binding site for one or more specific proteins. The poly(A) tail and its associated proteins probably help protect mRNA from enzymatic destruction. Many prokaryotic mRNAs also acquire poly(A) tails, but these tails stimulate decay of mRNA rather than protecting it from degradation.

The poly(A) tail is added in a multistep process. The transcript is extended beyond the site where the poly(A) tail is to be added, then is cleaved at the poly(A) addition site by an endonuclease component of a large enzyme complex, again associated with the CTD of RNA polymerase II (Fig. 26-17). The mRNA site where cleavage occurs is marked by two sequence elements: the highly conserved sequence (5')AAUAAA(3'), 10 to 30

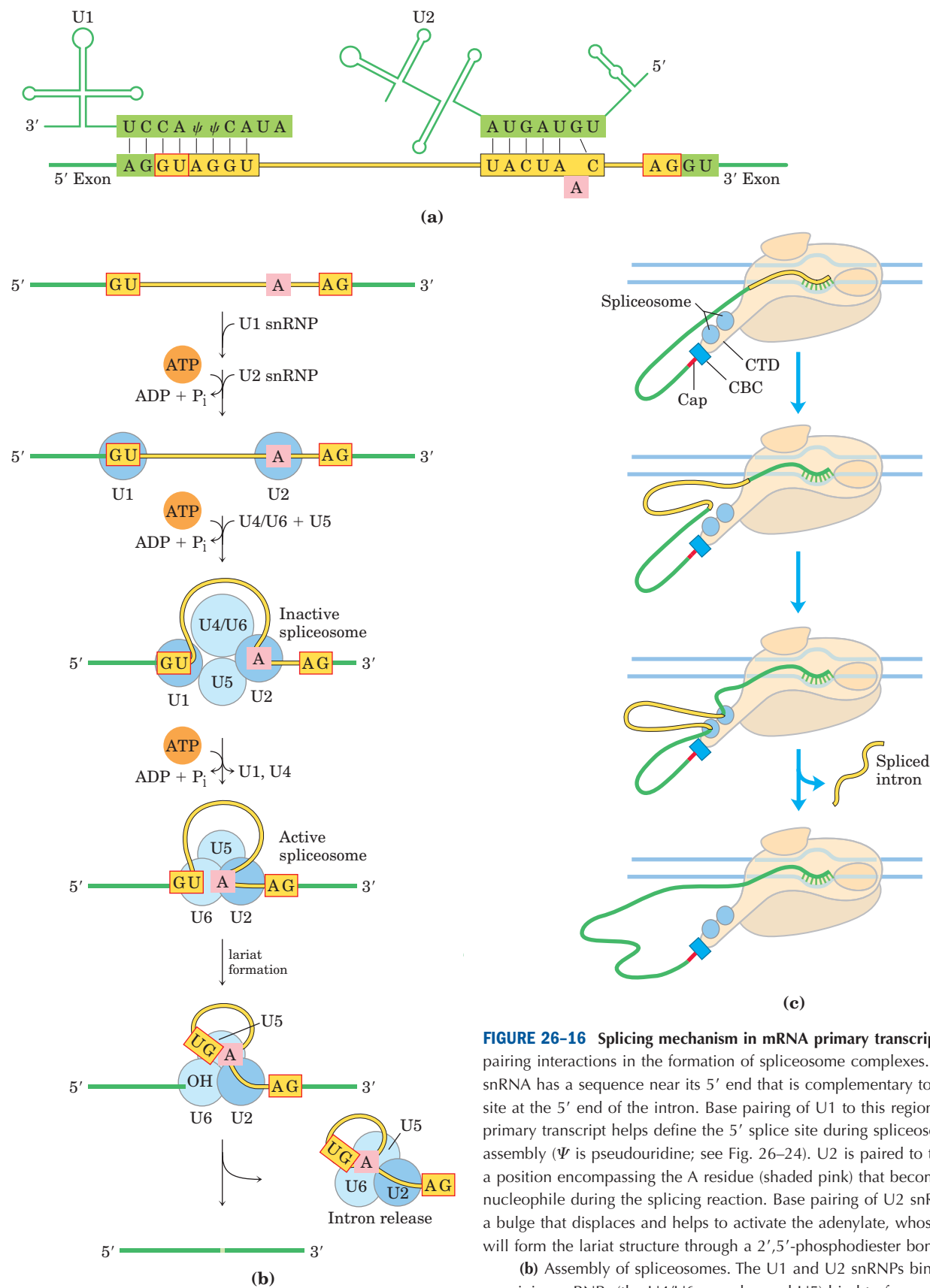


FIGURE 26-16 Splicing mechanism in mRNA primary transcripts. (a) RNA pairing interactions in the formation of spliceosome complexes. The U1 snRNA has a sequence near its 5' end that is complementary to the splice site at the 5' end of the intron. Base pairing of U1 to this region of the primary transcript helps define the 5' splice site during spliceosome assembly (Ψ is pseudouridine; see Fig. 26-24). U2 is paired to the intron at a position encompassing the A residue (shaded pink) that becomes the nucleophile during the splicing reaction. Base pairing of U2 snRNA causes a bulge that displaces and helps to activate the adenylate, whose 2' OH will form the lariat structure through a 2',5'-phosphodiester bond.

(b) Assembly of spliceosomes. The U1 and U2 snRNPs bind, then the remaining snRNPs (the U4/U6 complex and U5) bind to form an inactive spliceosome. Internal rearrangements convert this species to an active spliceosome in which U1 and U4 have been expelled and U6 is paired with both the 5' splice site and U2. This is followed by the catalytic steps, which parallel those of the splicing of group II introns (see Fig. 26-15).

(c) Coordination of splicing with transcription provides an attractive mechanism for bringing the two splice sites together. See the text for details.

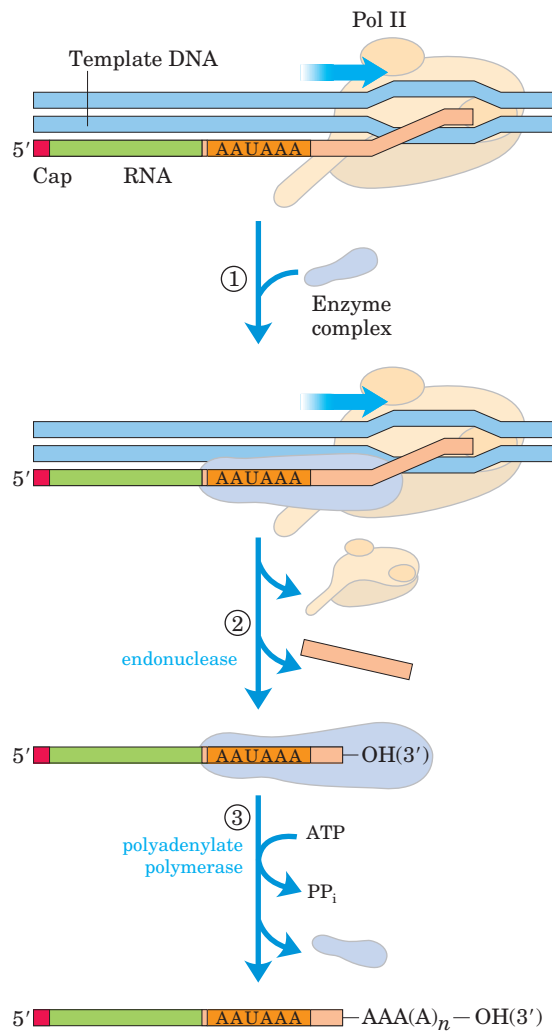


FIGURE 26-17 Addition of the poly(A) tail to the primary RNA transcript of eukaryotes. Pol II synthesizes RNA beyond the segment of the transcript containing the cleavage signal sequences, including the highly conserved upstream sequence (5')AAUAAA. ① The cleavage signal sequence is bound by an enzyme complex that includes an endonuclease, a polyadenylate polymerase, and several other multisubunit proteins involved in sequence recognition, stimulation of cleavage, and regulation of the length of the poly(A) tail. ② The RNA is cleaved by the endonuclease at a point 10 to 30 nucleotides 3' to (downstream of) the sequence AAUAAA. ③ The polyadenylate polymerase synthesizes a poly(A) tail 80 to 250 nucleotides long, beginning at the cleavage site.

nucleotides on the 5' side (upstream) of the cleavage site, and a less well-defined sequence rich in G and U residues, 20 to 40 nucleotides downstream of the cleavage site. Cleavage generates the free 3'-hydroxyl group that defines the end of the mRNA, to which A residues are immediately added by **polyadenylate polymerase**, which catalyzes the reaction



where $n = 80$ to 250 . This enzyme does not require a template but does require the cleaved mRNA as a primer.

The overall processing of a typical eukaryotic mRNA is summarized in Figure 26-18. In some cases the polypeptide-coding region of the mRNA is also modified by RNA “editing” (see Box 27-1 for details). This editing includes processes that add or delete bases in the coding regions

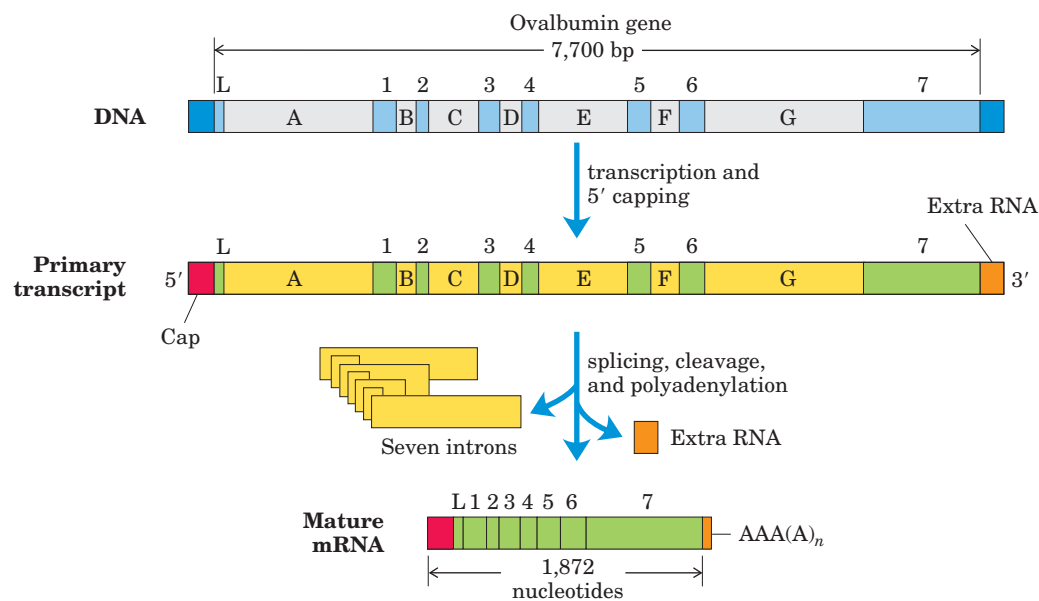


FIGURE 26-18 Overview of the processing of a eukaryotic mRNA. The ovalbumin gene, shown here, has introns A to G and exons 1 to 7 and L (L encodes a signal peptide sequence that targets the protein for export from the cell; see Fig. 27-34). About three-quarters of the

RNA is removed during processing. Pol II extends the primary transcript well beyond the cleavage and polyadenylation site (“extra RNA”) before terminating transcription. Termination signals for Pol II have not yet been defined.

of primary transcripts or that change the sequence (by, for example, enzymatic deamination of a C residue to create a U residue). A particularly dramatic example occurs in trypanosomes, which are parasitic protozoa: large regions of an mRNA are synthesized without any uridylate, and the U residues are inserted later by RNA editing.

A Gene Can Give Rise to Multiple Products by Differential RNA Processing

The transcription of introns seems to consume cellular resources and energy without returning any benefit to the organism, but introns may confer an advantage not yet fully appreciated by scientists. Introns may be vestiges of a molecular parasite not unlike transposons (Chapter 25). Although the benefits of introns are not yet clear in most cases, cells have evolved to take advantage of the splicing pathways to alter the expression of certain genes.

Most eukaryotic mRNA transcripts produce only one mature mRNA and one corresponding polypeptide, but some can be processed in more than one way to produce *different* mRNAs and thus different polypeptides. The primary transcript contains molecular signals for all the alternative processing pathways, and the pathway favored in a given cell is determined by processing factors, RNA-binding proteins that promote one particular path.

Complex transcripts can have either more than one site for cleavage and polyadenylation or alternative splicing patterns, or both. If there are two or more sites for cleavage and polyadenylation, use of the one closest to the 5' end will remove more of the primary transcript sequence (Fig. 26–19a). This mechanism, called poly(A) site choice, generates diversity in the variable domains of immunoglobulin heavy chains. Alternative splicing patterns (Fig. 26–19b) produce, from a common primary transcript, three different forms of the myosin heavy chain at different stages of fruit fly development. *Both* mechanisms come into play when a single RNA transcript is processed differently to produce two different hormones: the calcium-regulating hormone calcitonin in rat thyroid and calcitonin-gene-related peptide (CGRP) in rat brain (Fig. 26–20).

Ribosomal RNAs and tRNAs Also Undergo Processing

Posttranscriptional processing is not limited to mRNA. Ribosomal RNAs of both prokaryotic and eukaryotic cells are made from longer precursors called **preribosomal RNAs**, or pre-rRNAs, synthesized by Pol I. In bacteria, 16S, 23S, and 5S rRNAs (and some tRNAs, although most tRNAs are encoded elsewhere) arise from a single 30S RNA precursor of about 6,500 nucleotides. RNA at both ends of the 30S precursor and segments between the rRNAs are removed during processing (Fig. 26–21).

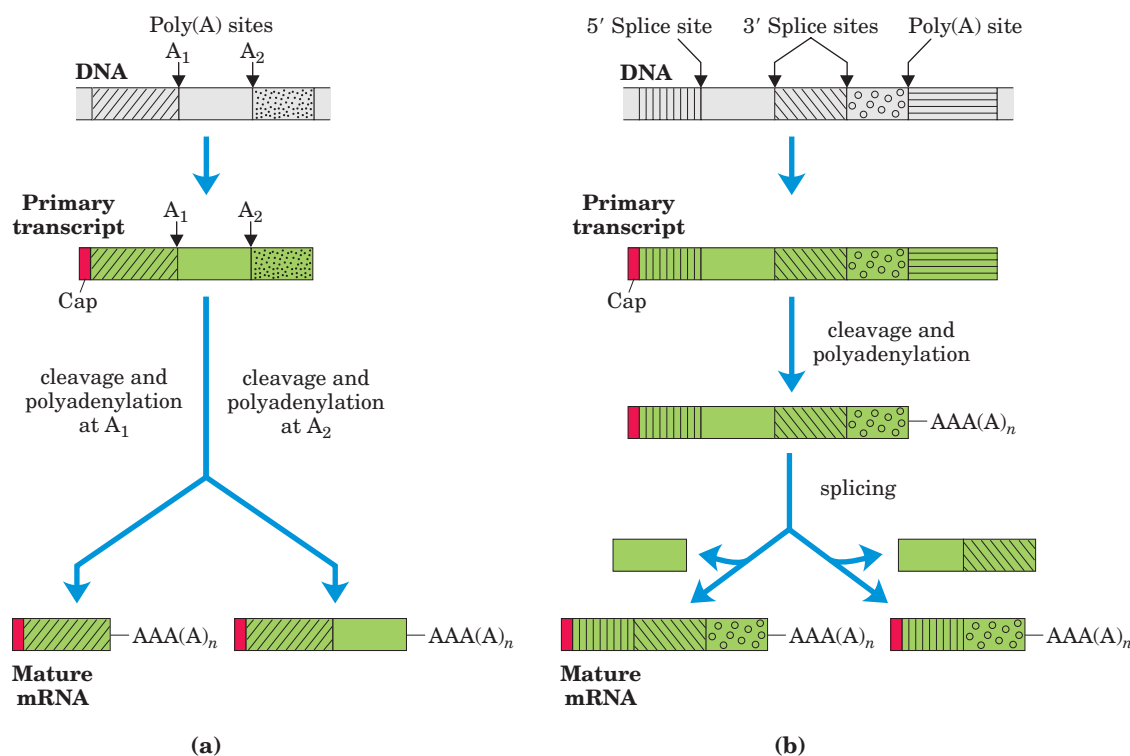


FIGURE 26-19 Two mechanisms for the alternative processing of complex transcripts in eukaryotes. (a) Alternative cleavage and polyadenylation patterns. Two poly(A) sites, A₁ and A₂, are shown.

(b) Alternative splicing patterns. Two different 3' splice sites are shown. In both mechanisms, different mature mRNAs are produced from the same primary transcript.

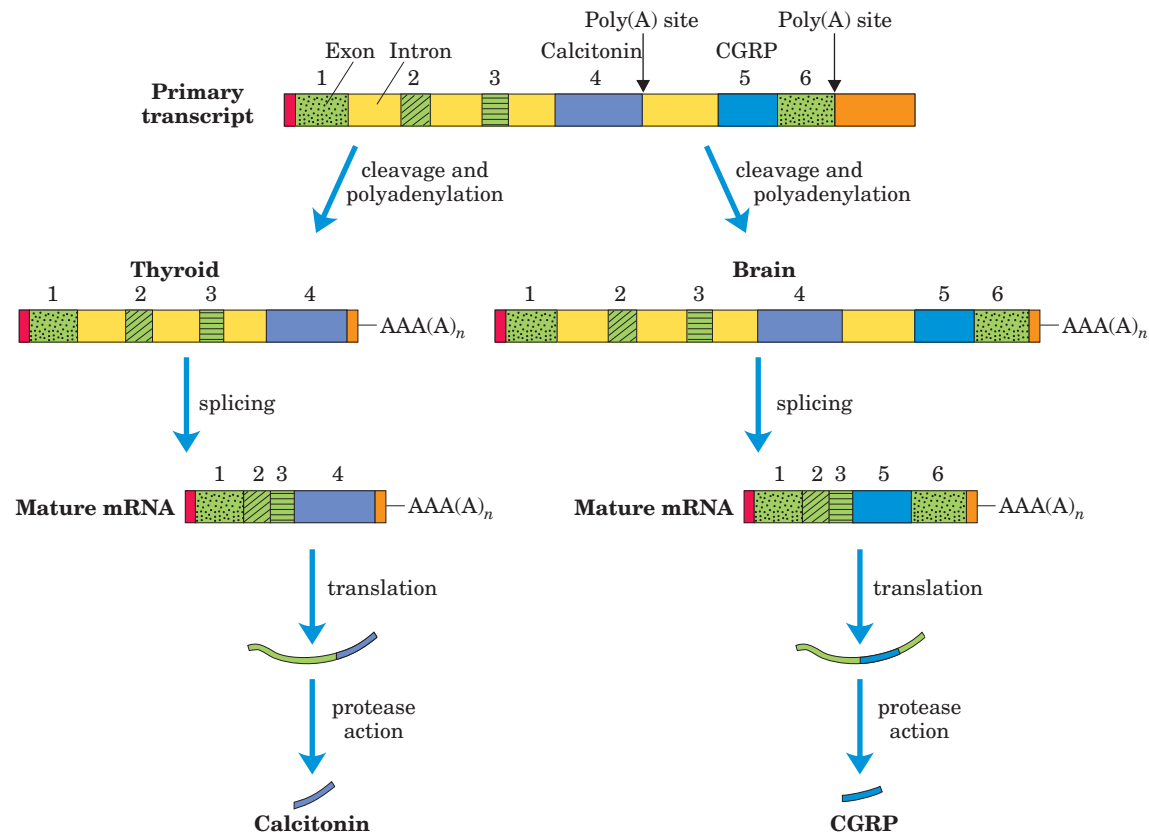


FIGURE 26-20 Alternative processing of the calcitonin gene transcript in rats. The primary transcript has two poly(A) sites; one predominates in the brain, the other in the thyroid. In the brain, splicing eliminates the calcitonin exon (exon 4); in the thyroid, this exon is re-

tained. The resulting peptides are processed further to yield the final hormone products: calcitonin-gene-related peptide (CGRP) in the brain and calcitonin in the thyroid.

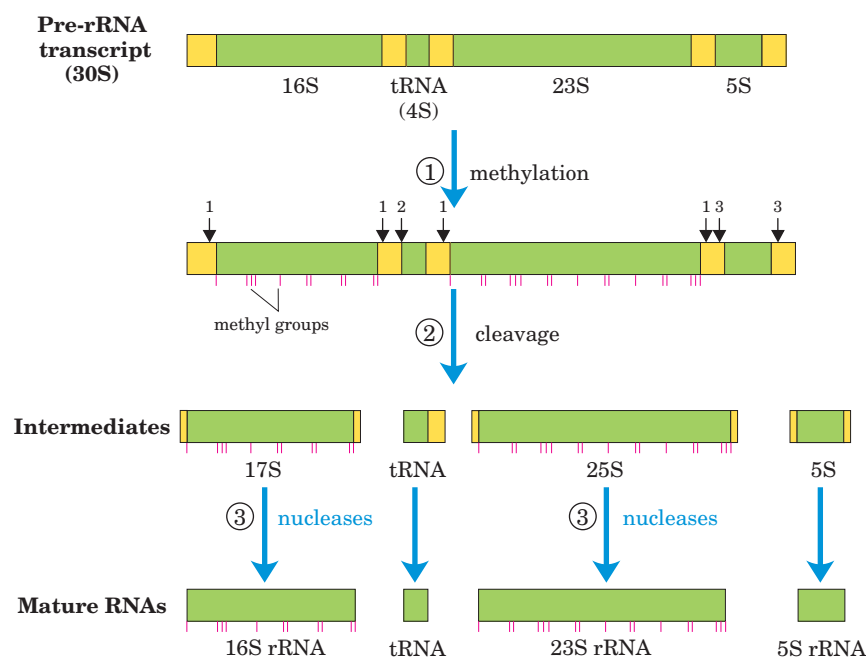


FIGURE 26-21 Processing of pre-rRNA transcripts in bacteria. ① Before cleavage, the 30S RNA precursor is methylated at specific bases. ② Cleavage liberates precursors of rRNAs and tRNA(s). Cleavage at the points labeled 1, 2, and 3 is carried out by the enzymes RNase III, RNase P, and RNase E, respectively. As discussed later in the text, RNase P is a ribozyme. ③ The final 16S, 23S, and 5S rRNA products result from the action of a variety of specific nucleases. The seven copies of the gene for pre-rRNA in the *E. coli* chromosome differ in the number, location, and identity of tRNAs included in the primary transcript. Some copies of the gene have additional tRNA gene segments between the 16S and 23S rRNA segments and at the far 3' end of the primary transcript.

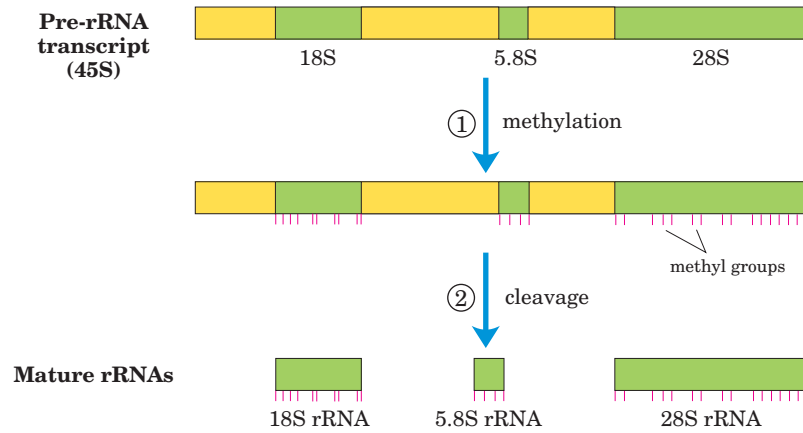


FIGURE 26-22 Processing of pre-rRNA transcripts in vertebrates. In step ①, the 45S precursor is methylated at more than 100 of its 14,000 nucleotides, mostly on the 2'-OH groups of ribose units retained in the final products. ② A series of enzymatic cleavages produces the 18S, 5.8S, and 28S rRNAs. The cleavage reactions require RNAs found in the nucleolus, called small nucleolar RNAs (snoRNAs), within protein complexes reminiscent of spliceosomes. The 5S rRNA is produced separately.

The genome of *E. coli* encodes seven pre-rRNA molecules. All these genes have essentially identical rRNA-coding regions, but they differ in the segments between these regions. The segment between the 16S and 23S rRNA genes generally encodes one or two tRNAs, with different tRNAs arising from different pre-rRNA transcripts. Coding sequences for tRNAs are also found on the 3' side of the 5S rRNA in some precursor transcripts.

In eukaryotes, a 45S pre-rRNA transcript is processed in the nucleolus to form the 18S, 28S, and 5.8S rRNAs characteristic of eukaryotic ribosomes (Fig. 26-22). The 5S rRNA of most eukaryotes is made as a completely separate transcript by a different polymerase (Pol III instead of Pol I).

Most cells have 40 to 50 distinct tRNAs, and eukaryotic cells have multiple copies of many of the tRNA

genes. Transfer RNAs are derived from longer RNA precursors by enzymatic removal of nucleotides from the 5' and 3' ends (Fig. 26-23). In eukaryotes, introns are present in a few tRNA transcripts and must be excised. Where two or more different tRNAs are contained in a single primary transcript, they are separated by enzymatic cleavage. The endonuclease RNase P, found in all organisms, removes RNA at the 5' end of tRNAs. This enzyme contains both protein and RNA. The RNA component is essential for activity, and in bacterial cells it can carry out its processing function with precision even without the protein component. RNase P is therefore another example of a catalytic RNA, as described in more detail below. The 3' end of tRNAs is processed by one or more nucleases, including the exonuclease RNase D.

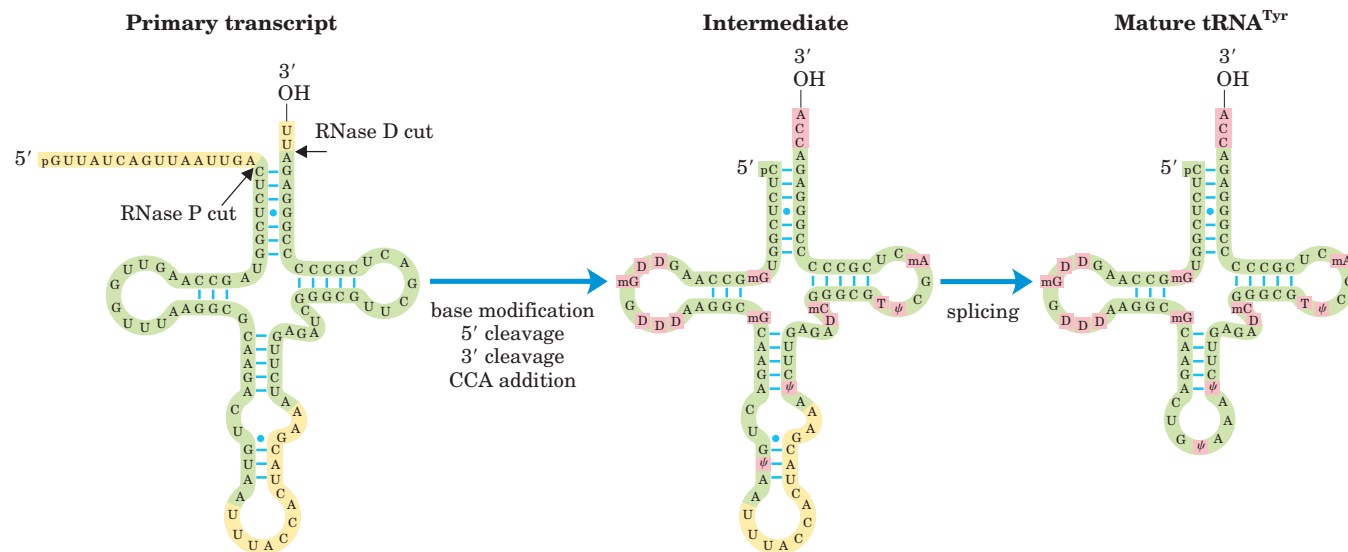


FIGURE 26-23 Processing of tRNAs in bacteria and eukaryotes. The yeast tRNA^{Tyr} (the tRNA specific for tyrosine binding; see Chapter 27) is used to illustrate the important steps. The nucleotide sequences shown in yellow are removed from the primary transcript. The ends are processed first, the 5' end before the 3' end. CCA is then added to the 3' end, a necessary step in processing eukaryotic tRNAs and

those bacterial tRNAs that lack this sequence in the primary transcript. While the ends are being processed, specific bases in the rest of the transcript are modified (see Fig. 26-24). For the eukaryotic tRNA shown here, the final step is splicing of the 14-nucleotide intron. Introns are found in some eukaryotic tRNAs but not in bacterial tRNAs.

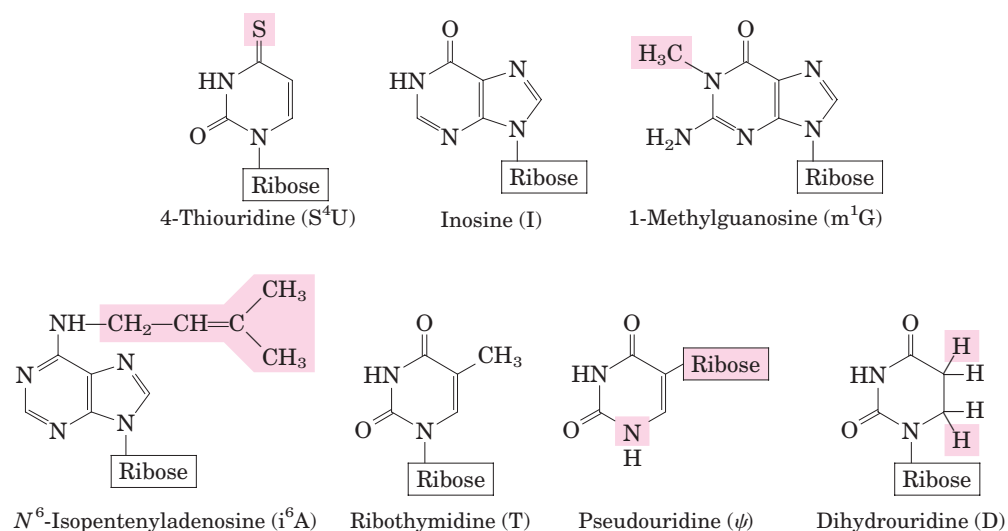


FIGURE 26-24 Some modified bases of tRNAs, produced in posttranscriptional reactions.

The standard symbols (used in Fig. 26-23) are shown in parentheses. Note the unusual ribose attachment point in pseudouridine.

Transfer RNA precursors may undergo further posttranscriptional processing. The 3'-terminal trinucleotide CCA(3') to which an amino acid will be attached during protein synthesis (Chapter 27) is absent from some bacterial and all eukaryotic tRNA precursors and is added during processing (Fig. 26-23). This addition is carried out by tRNA nucleotidyltransferase, an unusual enzyme that binds the three ribonucleoside triphosphate precursors in separate active sites and catalyzes formation of the phosphodiester bonds to produce the CCA(3') sequence. The creation of this defined sequence of nucleotides is therefore not dependent on a DNA or RNA template—the template is the binding site of the enzyme.

The final type of tRNA processing is the modification of some of the bases by methylation, deamination, or reduction (Fig. 26-24). In the case of pseudouridine (Ψ), the base (uracil) is removed and reattached to the sugar through C-5. Some of these modified bases occur at characteristic positions in all tRNAs (Fig. 26-23).

RNA Enzymes Are the Catalysts of Some Events in RNA Metabolism

The study of posttranscriptional processing of RNA molecules led to one of the most exciting discoveries in modern biochemistry—the existence of RNA enzymes. The best-characterized ribozymes are the self-splicing group I introns, RNase P, and the hammerhead ribozyme (discussed below). Most of the activities of these ribozymes are based on two fundamental reactions: transesterification (Fig. 26-13) and phosphodiester bond hydrolysis (cleavage). The substrate for ribozymes is often an RNA molecule, and it may even be part of the ribozyme itself. When its substrate is RNA, an RNA cat-

alyst can make use of base-pairing interactions to align the substrate for the reaction.

Ribozymes vary greatly in size. A self-splicing group I intron may have more than 400 nucleotides. The hammerhead ribozyme consists of two RNA strands with only 41 nucleotides in all (Fig. 26-25). As with protein enzymes, the three-dimensional structure of ribozymes is important for function. Ribozymes are inactivated by heating above their melting temperature or by addition of denaturing agents or complementary oligonucleotides, which disrupt normal base-pairing patterns. Ribozymes can also be inactivated if essential nucleotides are changed. The secondary structure of a self-splicing group I intron from the 26S rRNA precursor of *Tetrahymena* is shown in detail in Figure 26-26.

Enzymatic Properties of Group I Introns Self-splicing group I introns share several properties with enzymes besides accelerating the reaction rate, including their kinetic behaviors and their specificity. Binding of the guanosine cofactor (Fig. 26-13) to the *Tetrahymena* group I rRNA intron (Fig. 26-26) is saturable ($K_m \approx 30 \mu M$) and can be competitively inhibited by 3'-deoxyguanosine. The intron is very precise in its excision reaction, largely due to a segment called the **internal guide sequence** that can base-pair with exon sequences near the 5' splice site (Fig. 26-26). This pairing promotes the alignment of specific bonds to be cleaved and rejoined.

Because the intron itself is chemically altered during the splicing reaction—its ends are cleaved—it may appear to lack one key enzymatic property: the ability to catalyze multiple reactions. Closer inspection has shown that after excision, the 414 nucleotide intron from *Tetrahymena* rRNA can, in vitro, act as a true enzyme (but in vivo it is quickly degraded). A series of

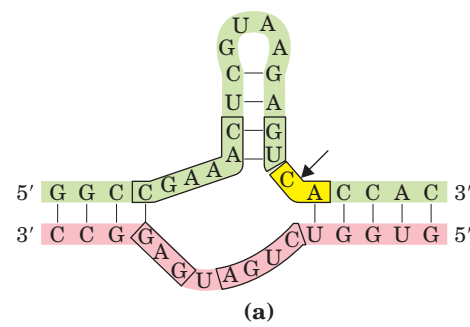


FIGURE 26-25 Hammerhead ribozyme. Certain viruslike elements called virusoids have small RNA genomes and usually require another virus to assist in their replication and/or packaging. Some virusoid RNAs include small segments that promote site-specific RNA cleavage reactions associated with replication. These segments are called hammerhead ribozymes, because their secondary structures are shaped like the head of a hammer. Hammerhead ribozymes have been defined and studied separately from the much larger viral RNAs. (a) The minimal sequences required for catalysis by the ribozyme. The boxed nucleotides are highly conserved and are required for catalytic function. The arrow indicates the site of self-cleavage. (b) Three-dimensional structure (PDB 1D 1MME). The strands are colored as in (a). The hammerhead ribozyme is a metalloenzyme; Mg^{2+} ions are required for activity. The phosphodiester bond at the site of self-cleavage is indicated by an arrow.

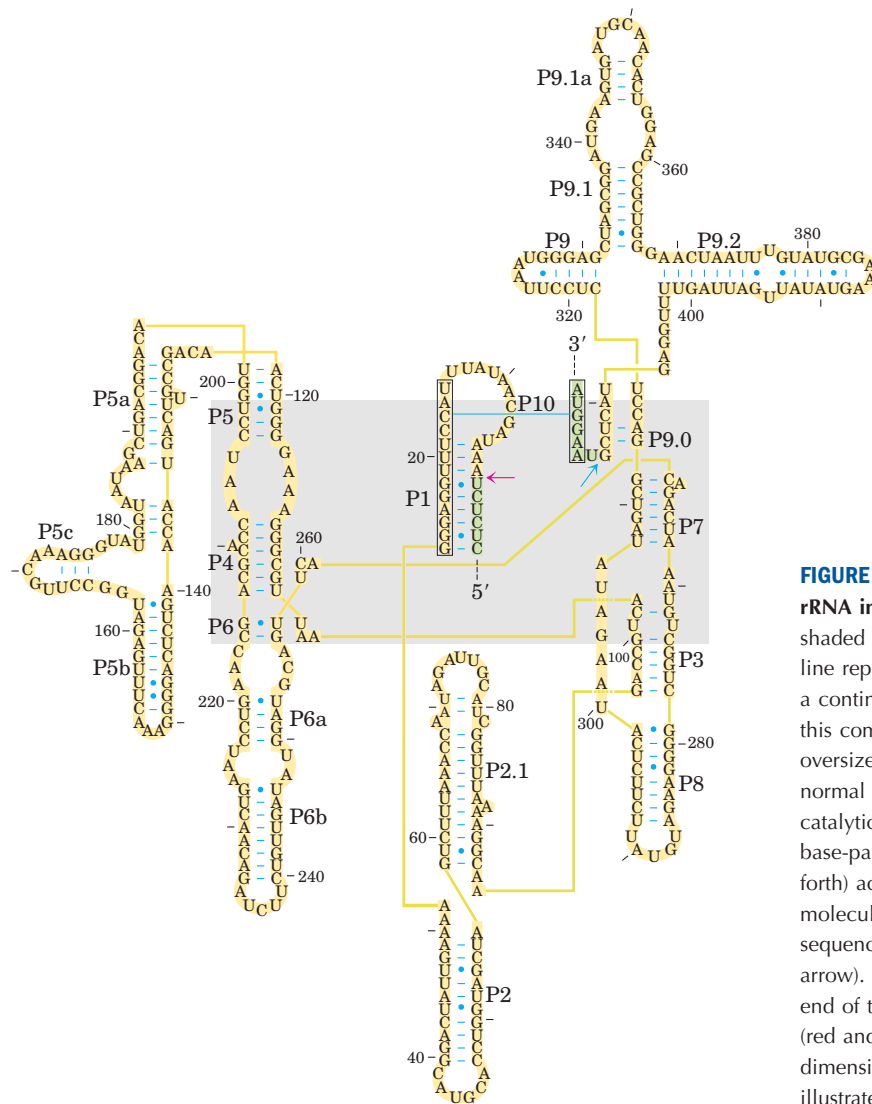
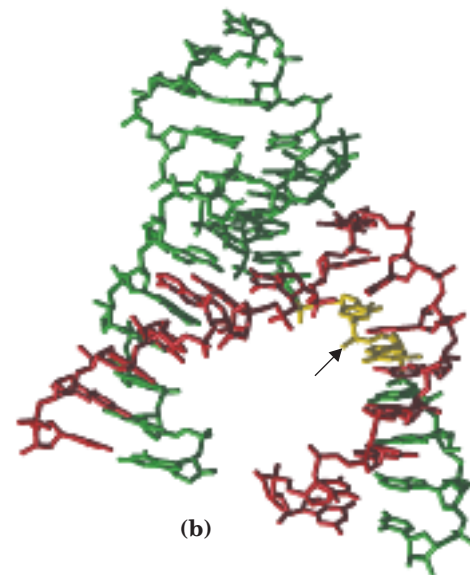


FIGURE 26-26 Secondary structure of the self-splicing rRNA intron from *Tetrahymena*. Intron sequences are shaded yellow, exon sequences green. Each thick yellow line represents a bond between neighboring nucleotides in a continuous sequence (a device necessitated by showing this complex molecule in two dimensions; similarly an oversize blue line between a C and G residue indicates normal base pairing); all nucleotides are shown. The catalytic core of the self-splicing activity is shaded. Some base-paired regions are labeled (P1, P3, P2.1, P5a, and so forth) according to an established convention for this RNA molecule. The P1 region, which contains the internal guide sequence (boxed), is the location of the 5' splice site (red arrow). Part of the internal guide sequence pairs with the end of the 3' exon, bringing the 5' and 3' splice sites (red and blue arrows) into close proximity. The three-dimensional structure of a large segment of this intron is illustrated in Figure 8-28c.

intramolecular cyclization and cleavage reactions in the excised intron leads to the loss of 19 nucleotides from its 5' end. The remaining 395 nucleotide, linear RNA—referred to as L-19 IVS—promotes nucleotidyl transfer reactions in which some oligonucleotides are lengthened at the expense of others (Fig. 26–27). The best substrates are oligonucleotides, such as a synthetic $(C)_5$ oligomer, that can base-pair with the same guanylate-rich internal guide sequence that held the 5' exon in place for self-splicing.

The enzymatic activity of the L-19 IVS ribozyme results from a cycle of transesterification reactions mechanistically similar to self-splicing. Each ribozyme molecule can process about 100 substrate molecules per hour and is not altered in the reaction; therefore the intron acts as a catalyst. It follows Michaelis-Menten kinetics, is specific for RNA oligonucleotide substrates, and can be competitively inhibited. The k_{cat}/K_m (specificity constant) is $10^3 \text{ M}^{-1} \text{ s}^{-1}$, lower than that of many enzymes, but the ribozyme accelerates hydrolysis by a factor of 10^{10} relative to the uncatalyzed reaction. It makes use of substrate orientation, covalent catalysis, and metal-ion catalysis—strategies used by protein enzymes.

Characteristics of Other Ribozymes *E. coli* RNase P has both an RNA component (the M1 RNA, with 377 nu-

cleotides) and a protein component (M_r 17,500). In 1983 Sidney Altman and Norman Pace and their coworkers discovered that under some conditions, the M1 RNA alone is capable of catalysis, cleaving tRNA precursors at the correct position. The protein component apparently serves to stabilize the RNA or facilitate its function in vivo. The RNase P ribozyme recognizes the three-dimensional shape of its pre-tRNA substrate, along with the CCA sequence, and thus can cleave the 5' leaders from diverse tRNAs (Fig. 26–23).

The known catalytic repertoire of ribozymes continues to expand. Some virusoids, small RNAs associated with plant RNA viruses, include a structure that promotes a self-cleavage reaction; the hammerhead ribozyme illustrated in Figure 26–25 is in this class, catalyzing the hydrolysis of an internal phosphodiester bond. The splicing reaction that occurs in a spliceosome seems to rely on a catalytic center formed by the U2, U5, and U6 snRNAs (Fig. 26–16). And perhaps most important, an RNA component of ribosomes catalyzes the synthesis of proteins (Chapter 27).

Exploring catalytic RNAs has provided new insights into catalytic function in general and has important implications for our understanding of the origin and evolution of life on this planet, a topic discussed in Section 26.3.

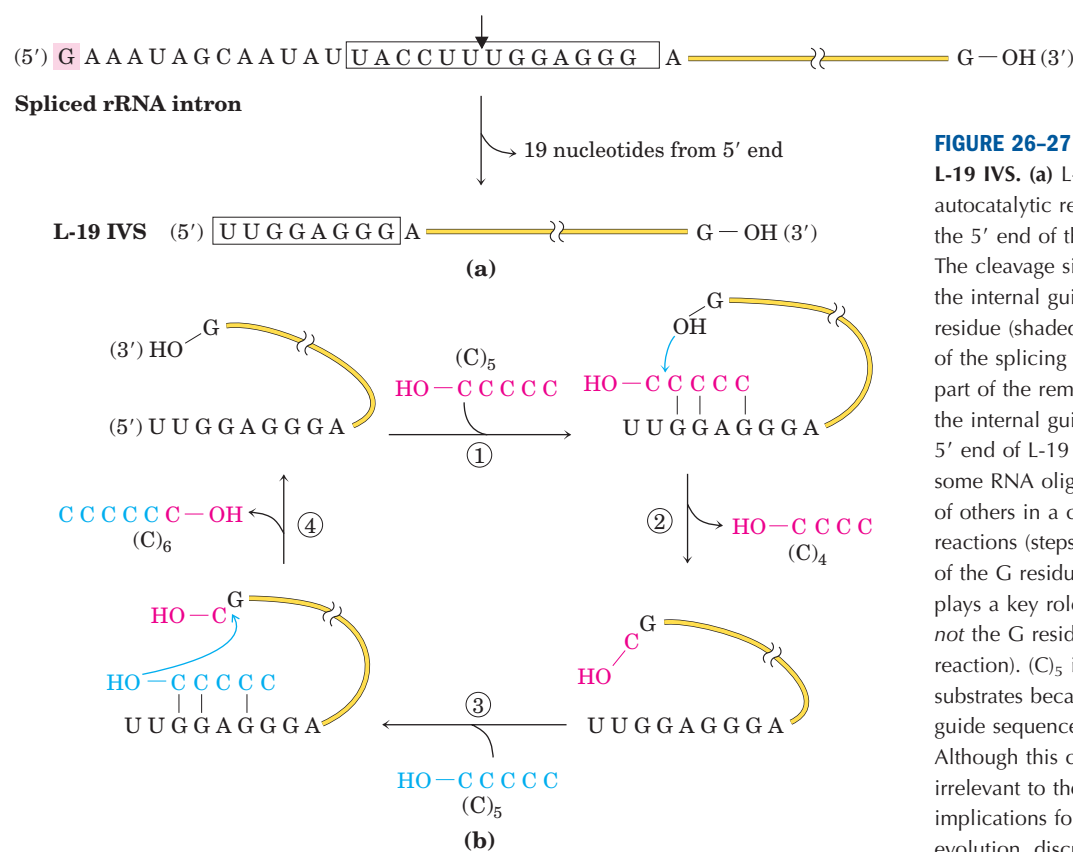



FIGURE 26–27 In vitro catalytic activity of L-19 IVS. (a) L-19 IVS is generated by the autocatalytic removal of 19 nucleotides from the 5' end of the spliced *Tetrahymena* intron. The cleavage site is indicated by the arrow in the internal guide sequence (boxed). The G residue (shaded pink) added in the first step of the splicing reaction (see Fig. 26–14) is part of the removed sequence. A portion of the internal guide sequence remains at the 5' end of L-19 IVS. (b) L-19 IVS lengthens some RNA oligonucleotides at the expense of others in a cycle of transesterification reactions (steps ① through ④). The 3' OH of the G residue at the 3' end of L-19 IVS plays a key role in this cycle (note that this is *not* the G residue added in the splicing reaction). $(C)_5$ is one of the ribozyme's better substrates because it can base-pair with the guide sequence remaining in the intron. Although this catalytic activity is probably irrelevant to the cell, it has important implications for current hypotheses on evolution, discussed at the end of this chapter.

Cellular mRNAs Are Degraded at Different Rates

The expression of genes is regulated at many levels. A crucial factor governing a gene's expression is the cellular concentration of its associated mRNA. The concentration of any molecule depends on two factors: its rate of synthesis and its rate of degradation. When synthesis and degradation of an mRNA are balanced, the concentration of the mRNA remains in a steady state. A change in either rate will lead to net accumulation or depletion of the mRNA. Degradative pathways ensure that mRNAs do not build up in the cell and direct the synthesis of unnecessary proteins.

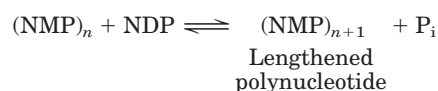
The rates of degradation vary greatly for mRNAs from different eukaryotic genes. For a gene product that is needed only briefly, the half-life of its mRNA may be only minutes or even seconds. Gene products needed constantly by the cell may have mRNAs that are stable over many cell generations. The average half-life of a vertebrate cell mRNA is about 3 hours, with the pool of each type of mRNA turning over about ten times per cell generation. The half-life of bacterial mRNAs is much shorter—only about 1.5 min—perhaps because of regulatory requirements.

Messenger RNA is degraded by ribonucleases present in all cells. In *E. coli*, the process begins with one or a few cuts by an endoribonuclease, followed by 3'→5' degradation by exoribonucleases. In lower eukaryotes, the major pathway involves first shortening the poly(A) tail, then decapping the 5' end and degrading the mRNA in the 5'→3' direction. A 3'→5' degradative pathway also exists and may be the major path in higher eukaryotes. All eukaryotes have a complex of up to ten conserved 3'→5' exoribonucleases, called the exosome, which is involved in the processing of the 3' end of rRNAs and tRNAs as well as the degradation of mRNAs.

A hairpin structure in bacterial mRNAs with a ρ -independent terminator (Fig. 26–7) confers stability against degradation. Similar hairpin structures can make some parts of a primary transcript more stable, leading to nonuniform degradation of transcripts. In eukaryotic cells, both the 3' poly(A) tail and the 5' cap are important to the stability of many mRNAs.  **Life Cycle of an mRNA**

Polynucleotide Phosphorylase Makes Random RNA-like Polymers

In 1955, Marianne Grunberg-Manago and Severo Ochoa discovered the bacterial enzyme **polynucleotide phosphorylase**, which in vitro catalyzes the reaction



Polynucleotide phosphorylase was the first nucleic acid-synthesizing enzyme discovered (Arthur Kornberg's discovery of DNA polymerase followed soon thereafter).

The reaction catalyzed by polynucleotide phosphorylase differs fundamentally from the polymerase activities discussed so far in that it is not template-dependent. The enzyme uses the 5'-diphosphates of ribonucleosides as substrates and cannot act on the homologous 5'-triphosphates or on deoxyribonucleoside 5'-diphosphates. The RNA polymer formed by polynucleotide phosphorylase contains the usual 3',5'-phosphodiester linkages, which can be hydrolyzed by ribonuclease. The reaction is readily reversible and can be pushed in the direction of breakdown of the polyribonucleotide by increasing the phosphate concentration. The probable function of this enzyme in the cell is the degradation of mRNAs to nucleoside diphosphates.

Because the polynucleotide phosphorylase reaction does not use a template, the polymer it forms does not have a specific base sequence. The reaction proceeds equally well with any or all of the four nucleoside diphosphates, and the base composition of the resulting polymer reflects nothing more than the relative concentrations of the 5'-diphosphate substrates in the medium.

Polynucleotide phosphorylase can be used in the laboratory to prepare RNA polymers with many different base sequences and frequencies. Synthetic RNA polymers of this sort were critical for deducing the genetic code for the amino acids (Chapter 27).

SUMMARY 26.2 RNA Processing

- Eukaryotic mRNAs are modified by addition of a 7-methylguanosine residue at the 5' end and by cleavage and polyadenylation at the 3' end to form a long poly(A) tail.
- Many primary mRNA transcripts contain introns (noncoding regions), which are removed by splicing. Excision of the group I introns found in some rRNAs requires a guanosine cofactor. Some group I and group II introns are capable of self-splicing; no protein enzymes are required. Nuclear mRNA precursors have a third class (the largest class) of introns, which are spliced



Marianne Grunberg-Manago



Severo Ochoa,
1905–1993

with the aid of RNA-protein complexes called snRNPs, assembled into spliceosomes. A fourth class of introns, found in some tRNAs, is the only class known to be spliced by protein enzymes.

- Ribosomal RNAs and transfer RNAs are derived from longer precursor RNAs, trimmed by nucleases. Some bases are modified enzymatically during the maturation process.
- The self-splicing introns and the RNA component of RNase P (which cleaves the 5' end of tRNA precursors) are two examples of ribozymes. These biological catalysts have the properties of true enzymes. They generally promote hydrolytic cleavage and transesterification, using RNA as substrate. Combinations of these reactions can be promoted by the excised group I intron of *Tetrahymena* rRNA, resulting in a type of RNA polymerization reaction.
- Polynucleotide phosphorylase reversibly forms RNA-like polymers from ribonucleoside 5'-diphosphates, adding or removing ribonucleotides at the 3'-hydroxyl end of the polymer. The enzyme degrades RNA in vivo.

26.3 RNA-Dependent Synthesis of RNA and DNA

In our discussion of DNA and RNA synthesis up to this point, the role of the template strand has been reserved for DNA. However, some enzymes use an RNA template for nucleic acid synthesis. With the very important exception of viruses with an RNA genome, these enzymes play only a modest role in information pathways. RNA viruses are the source of most RNA-dependent polymerases characterized so far.

The existence of RNA replication requires an elaboration of the central dogma (Fig. 26–28; contrast this with the diagram on p. 922). The enzymes involved in

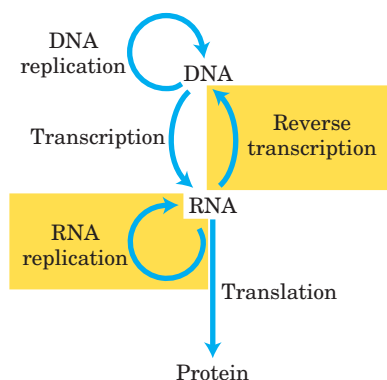


FIGURE 26-28 Extension of the central dogma to include RNA-dependent synthesis of RNA and DNA.

RNA replication have profound implications for investigations into the nature of self-replicating molecules that may have existed in prebiotic times.

Reverse Transcriptase Produces DNA from Viral RNA

Certain RNA viruses that infect animal cells carry within the viral particle an RNA-dependent DNA polymerase called **reverse transcriptase**. On infection, the single-stranded RNA viral genome (~10,000 nucleotides) and the enzyme enter the host cell. The reverse transcriptase first catalyzes the synthesis of a DNA strand complementary to the viral RNA (Fig. 26–29), then degrades the RNA strand of the viral RNA-DNA hybrid and replaces it with DNA. The resulting duplex DNA often becomes incorporated into the genome of the eukaryotic host cell. These integrated (and dormant) viral genes can be activated and transcribed, and the gene products—viral proteins and the viral RNA genome itself—packaged as new viruses. The RNA viruses that contain reverse transcriptases are known as **retroviruses** (*retro* is the Latin prefix for “backward”).

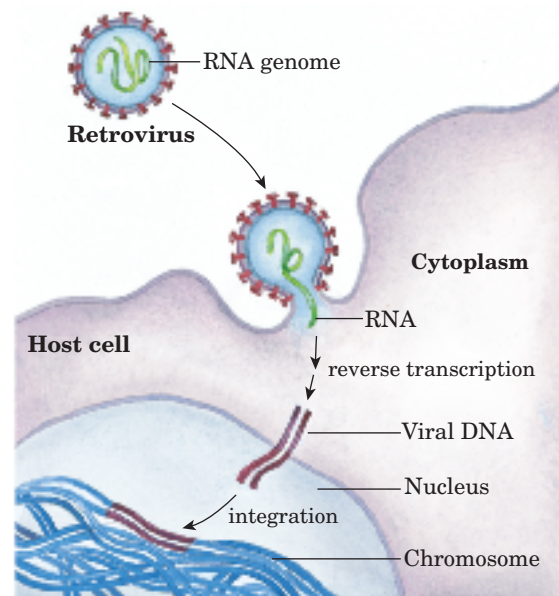


FIGURE 26-29 Retroviral infection of a mammalian cell and integration of the retrovirus into the host chromosome. Viral particles entering the host cell carry viral reverse transcriptase and a cellular tRNA (picked up from a former host cell) already base-paired to the viral RNA. The tRNA facilitates immediate conversion of viral RNA to double-stranded DNA by the action of reverse transcriptase, as described in the text. Once converted to double-stranded DNA, the DNA enters the nucleus and is integrated into the host genome. The integration is catalyzed by a virally encoded integrase. Integration of viral DNA into host DNA is mechanistically similar to the insertion of transposons in bacterial chromosomes (see Fig. 25–43). For example, a few base pairs of host DNA become duplicated at the site of integration, forming short repeats of 4 to 6 bp at each end of the inserted retroviral DNA (not shown).

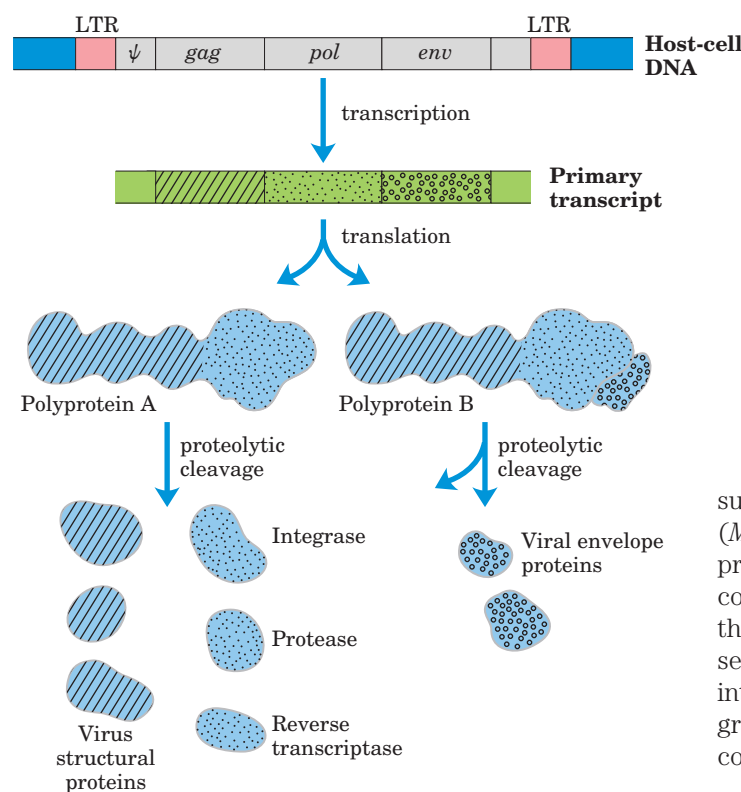


FIGURE 26-30 Structure and gene products of an integrated retroviral genome. The long terminal repeats (LTRs) have sequences needed for the regulation and initiation of transcription. The sequence denoted Ψ is required for packaging of retroviral RNAs into mature viral particles. Transcription of the retroviral DNA produces a primary transcript encompassing the *gag*, *pol*, and *env* genes. Translation (Chapter 27) produces a polyprotein, a single long polypeptide derived from the *gag* and *pol* genes, which is cleaved into six distinct proteins. Splicing of the primary transcript yields an mRNA derived largely from the *env* gene, which is also translated into a polyprotein, then cleaved to generate viral envelope proteins.

The existence of reverse transcriptases in RNA viruses was predicted by Howard Temin in 1962, and the enzymes were ultimately detected by Temin and, independently, by David Baltimore in 1970. Their discovery aroused much attention as dogma-shaking proof that genetic information can flow “backward” from RNA to DNA.

Retroviruses typically have three genes: *gag* (derived from the historical designation *group associated antigen*), *pol*, and *env* (Fig. 26–30). The transcript that contains *gag* and *pol* is translated into a long “polyprotein,” a single large polypeptide that is cleaved into six proteins with distinct functions. The proteins derived from the *gag* gene make up the interior core of the viral particle. The *pol* gene encodes the protease that cleaves the long polypeptide, an integrase that inserts the viral DNA into the host chromosomes, and reverse transcriptase. Many reverse transcriptases have two



Howard Temin,
1934–1994



David Baltimore

subunits, α and β . The *pol* gene specifies the β subunit (M_r 90,000), and the α subunit (M_r 65,000) is simply a proteolytic fragment of the β subunit. The *env* gene encodes the proteins of the viral envelope. At each end of the linear RNA genome are long terminal repeat (LTR) sequences of a few hundred nucleotides. Transcribed into the duplex DNA, these sequences facilitate integration of the viral chromosome into the host DNA and contain promoters for viral gene expression.

Reverse transcriptases catalyze three different reactions: (1) RNA-dependent DNA synthesis, (2) RNA degradation, and (3) DNA-dependent DNA synthesis. Like many DNA and RNA polymerases, reverse transcriptases contain Zn^{2+} . Each transcriptase is most active with the RNA of its own virus, but each can be used experimentally to make DNA complementary to a variety of RNAs. The DNA and RNA synthesis and RNA degradation activities use separate active sites on the protein. For DNA synthesis to begin, the reverse transcriptase requires a primer, a cellular tRNA obtained during an earlier infection and carried within the viral particle. This tRNA is base-paired at its 3' end with a complementary sequence in the viral RNA. The new DNA strand is synthesized in the 5'→3' direction, as in all RNA and DNA polymerase reactions. Reverse transcriptases, like RNA polymerases, do not have 3'→5' proofreading exonucleases. They generally have error rates of about 1 per 20,000 nucleotides added. An error rate this high is extremely unusual in DNA replication and appears to be a feature of most enzymes that replicate the genomes of RNA viruses. A consequence is a higher mutation rate and faster rate of viral evolution, which is a factor in the frequent appearance of new strains of disease-causing retroviruses.

Reverse transcriptases have become important reagents in the study of DNA-RNA relationships and in DNA cloning techniques. They make possible the synthesis of DNA complementary to an mRNA template, and synthetic DNA prepared in this manner, called **complementary DNA (cDNA)**, can be used to clone cellular genes (see Fig. 9–14).



FIGURE 26–31 Rous sarcoma virus genome. The *src* gene encodes a tyrosine-specific protein kinase, one of a class of enzymes known to function in systems that affect cell division, cell-cell interactions, and intercellular communication (Chapter 12). The same gene is found in

chicken DNA (the usual host for this virus) and in the genomes of many other eukaryotes, including humans. When associated with the Rous sarcoma virus, this oncogene is often expressed at abnormally high levels, contributing to unregulated cell division and cancer.

Some Retroviruses Cause Cancer and AIDS

Retroviruses have featured prominently in recent advances in the molecular understanding of cancer. Most retroviruses do not kill their host cells but remain integrated in the cellular DNA, replicating when the cell divides. Some retroviruses, classified as RNA tumor viruses, contain an oncogene that can cause the cell to grow abnormally (see Fig. 12–47). The first retrovirus of this type to be studied was the Rous sarcoma virus (also called avian sarcoma virus; Fig. 26–31), named for F. Peyton Rous, who studied chicken tumors now known to be caused by this virus. Since the initial discovery of oncogenes by Harold Varmus and Michael Bishop, many dozens of such genes have been found in retroviruses.

The human immunodeficiency virus (HIV), which causes acquired immune deficiency syndrome (AIDS), is a retrovirus. Identified in 1983, HIV has an RNA genome with standard retroviral genes along with several other unusual genes (Fig. 26–32). Unlike many other retroviruses, HIV kills many of the cells it infects (principally T lymphocytes) rather than causing tumor formation. This gradually leads to suppression of the immune system in the host organism. The reverse transcriptase of HIV is even more error prone than other known reverse transcriptases—ten times more so—resulting in high mutation rates in this virus. One or more errors are generally made every time the viral genome is replicated, so any two viral RNA molecules are likely to differ.

Many modern vaccines for viral infections consist of one or more coat proteins of the virus, produced by methods described in Chapter 9. These proteins are not

infectious on their own but stimulate the immune system to recognize and resist subsequent viral invasions (Chapter 5). Because of the high error rate of the HIV reverse transcriptase, the *env* gene in this virus (along with the rest of the genome) undergoes very rapid mutation, complicating the development of an effective vaccine. However, repeated cycles of cell invasion and replication are needed to propagate an HIV infection, so inhibition of viral enzymes offers promise as an effective therapy. The HIV protease is targeted by a class of drugs called protease inhibitors (see Box 6–3). Reverse transcriptase is the target of some additional drugs widely used to treat HIV-infected individuals (Box 26–2).

Many Transposons, Retroviruses, and Introns May Have a Common Evolutionary Origin

Some well-characterized eukaryotic DNA transposons from sources as diverse as yeast and fruit flies have a structure very similar to that of retroviruses; these are sometimes called retrotransposons (Fig. 26–33). Retrotransposons encode an enzyme homologous to the retroviral reverse transcriptase, and their coding regions are flanked by LTR sequences. They transpose from one position to another in the cellular genome by means of an RNA intermediate, using reverse transcriptase to make a DNA copy of the RNA, followed by integration of the DNA at a new site. Most transposons in eukaryotes use this mechanism for transposition, distinguishing them from bacterial transposons, which move as DNA directly from one chromosomal location to another (see Fig. 25–43).

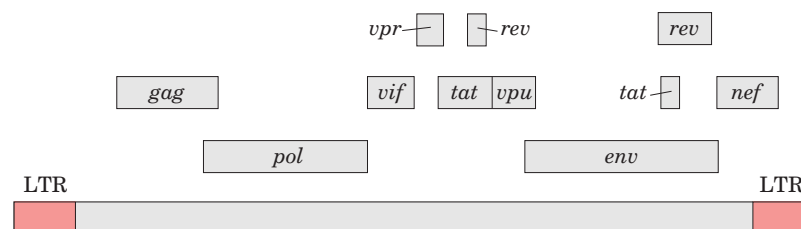


FIGURE 26–32 The genome of HIV, the virus that causes AIDS. In addition to the typical retroviral genes, HIV contains several small genes with a variety of functions (not identified here, and not all

known). Some of these genes overlap (see Box 27–1). Alternative splicing mechanisms produce many different proteins from this small (9.7×10^3 nucleotides) genome.

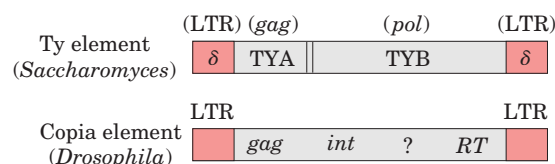


FIGURE 26-33 Eukaryotic transposons. The Ty element of the yeast *Saccharomyces* and the copia element of the fruit fly *Drosophila* serve as examples of eukaryotic transposons, which often have a structure similar to retroviruses but lack the *env* gene. The δ sequences of the Ty element are functionally equivalent to retroviral LTRs. In the copia element, *int* and *RT* are homologous to the integrase and reverse transcriptase segments, respectively, of the *pol* gene.

Retrotransposons lack an *env* gene and so cannot form viral particles. They can be thought of as defective viruses, trapped in cells. Comparisons between retroviruses and eukaryotic transposons suggest that reverse transcriptase is an ancient enzyme that predates the evolution of multicellular organisms.

Interestingly, many group I and group II introns are also mobile genetic elements. In addition to their self-splicing activities, they encode DNA endonucleases that promote their movement. During genetic exchanges between cells of the same species, or when DNA is introduced into a cell by parasites or by other means, these endonucleases promote insertion of the intron into an identical site in another DNA copy of a homologous gene that does not contain the intron, in a process termed **homing** (Fig. 26-34). Whereas group I intron homing is DNA-based, group II intron homing occurs through an RNA intermediate. The endonucleases of the group II introns have associated reverse transcriptase activity. The proteins can form complexes with the intron RNAs themselves, after the introns are spliced from the primary transcripts. Because the homing process involves insertion of the RNA intron into DNA and reverse transcription of the intron, the movement of these introns has been called retrohoming. Over time, every copy of a particular gene in a population may acquire the intron.



BOX 26-2 BIOCHEMISTRY IN MEDICINE

Fighting AIDS with Inhibitors of HIV Reverse Transcriptase

Research into the chemistry of template-dependent nucleic acid biosynthesis, combined with modern techniques of molecular biology, has elucidated the life cycle and structure of the human immunodeficiency virus, the retrovirus that causes AIDS. A few years after the isolation of HIV, this research resulted in the development of drugs capable of prolonging the lives of people infected by HIV.

The first drug to be approved for clinical use was AZT, a structural analog of deoxythymidine. AZT was first synthesized in 1964 by Jerome P. Horwitz. It failed as an anticancer drug (the purpose for which it was made), but in 1985 it was found to be a useful treatment for AIDS. AZT is taken up by T lymphocytes, immune system cells that are particularly vulnerable

to HIV infection, and converted to AZT triphosphate. (AZT triphosphate taken directly would be ineffective, because it cannot cross the plasma membrane.) HIV's reverse transcriptase has a higher affinity for AZT triphosphate than for dTTP, and binding of AZT triphosphate to this enzyme competitively inhibits dTTP binding. When AZT is added to the 3' end of the growing DNA strand, lack of a 3' hydroxyl means that the DNA strand is terminated prematurely and viral DNA synthesis grinds to a halt.

AZT triphosphate is not as toxic to the T lymphocytes themselves, because *cellular* DNA polymerases have a lower affinity for this compound than for dTTP. At concentrations of 1 to 5 μM , AZT affects HIV reverse transcription but not most cellular DNA replication. Unfortunately, AZT appears to be toxic to the bone marrow cells that are the progenitors of erythrocytes, and many individuals taking AZT develop anemia. AZT can increase the survival time of people with advanced AIDS by about a year, and it delays the onset of AIDS in those who are still in the early stages of HIV infection. Some other AIDS drugs, such as dideoxyinosine (DDI), have a similar mechanism of action. Newer drugs target and inactivate the HIV protease. Because of the high error rate of HIV reverse transcriptase and the resulting rapid evolution of HIV, the most effective treatments of HIV infections use a combination of drugs directed at both the protease and the reverse transcriptase.

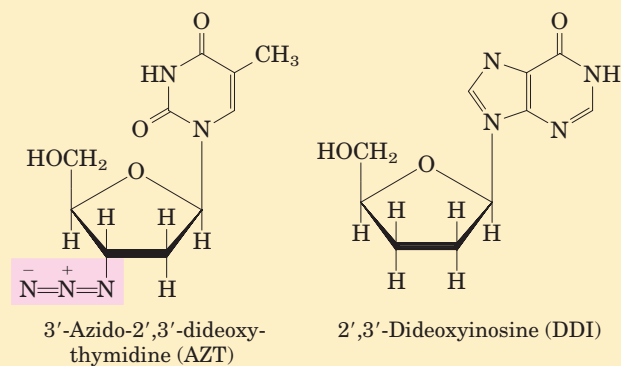


FIGURE 26-34 Introns that move: homing and retrohoming. Certain introns include a gene (shown in red) for enzymes that promote homing (type I introns) or retrohoming (type II introns). **(a)** The gene within the spliced intron is bound by a ribosome and translated. Type I homing introns specify a site-specific endonuclease, called a homing endonuclease. Type II retrohoming introns specify a protein with both endonuclease and reverse transcriptase activities.

(b) Homing. Allele *a* of a gene *X* containing a type I homing intron is present in a cell containing allele *b* of the same gene, which lacks the intron. The homing endonuclease produced by *a* cleaves *b* at the position corresponding to the intron in *a*, and double-strand break repair (recombination with allele *a*; see Fig. 25-31a) then creates a new copy of the intron in *b*. **(c)** Retrohoming. Allele *a* of gene *Y* contains a retrohoming type II intron; allele *b* lacks the intron. The spliced intron inserts itself into the coding strand of *b* in a reaction that is the reverse of the splicing that excised the intron from the primary transcript (see Fig. 26-15), except that here the insertion is into DNA rather than RNA. The noncoding DNA strand of *b* is then cleaved by the intron-encoded endonuclease/reverse transcriptase. This same enzyme uses the inserted RNA as a template to synthesize a complementary DNA strand. The RNA is then degraded by cellular ribonucleases and replaced with DNA.

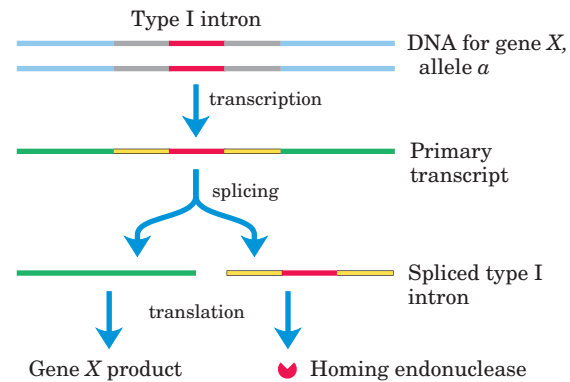
Much more rarely, the intron may insert itself into a new location in an unrelated gene. If this event does not kill the host cell, it can lead to the evolution and distribution of an intron in a new location. The structures and mechanisms used by mobile introns support the idea that at least some introns originated as molecular parasites whose evolutionary past can be traced to retroviruses and transposons.

Telomerase Is a Specialized Reverse Transcriptase

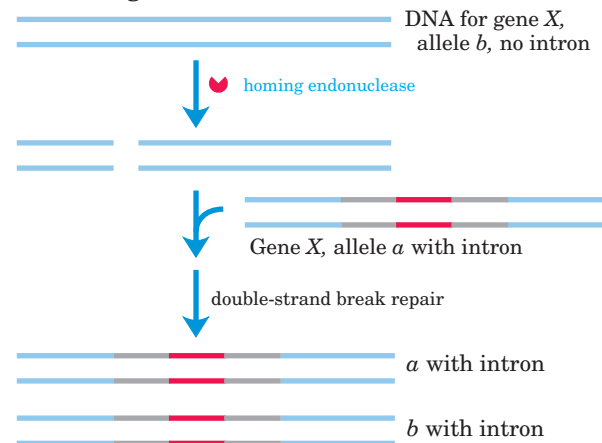
Telomeres, the structures at the ends of linear eukaryotic chromosomes (see Fig. 24-9), generally consist of many tandem copies of a short oligonucleotide sequence. This sequence usually has the form T_xG_y in one strand and C_yA_x in the complementary strand, where x and y are typically in the range of 1 to 4 (p. 930). Telomeres vary in length from a few dozen base pairs in some ciliated protozoans to tens of thousands of base pairs in mammals. The TG strand is longer than its complement, leaving a region of single-stranded DNA of up to a few hundred nucleotides at the 3' end.

The ends of a linear chromosome are not readily replicated by cellular DNA polymerases. DNA replication requires a template and primer, and beyond the end of a linear DNA molecule no template is available for the pairing of an RNA primer. Without a special mechanism for replicating the ends, chromosomes would be shortened somewhat in each cell generation. The enzyme **telomerase** solves this problem by adding telomeres to chromosome ends.

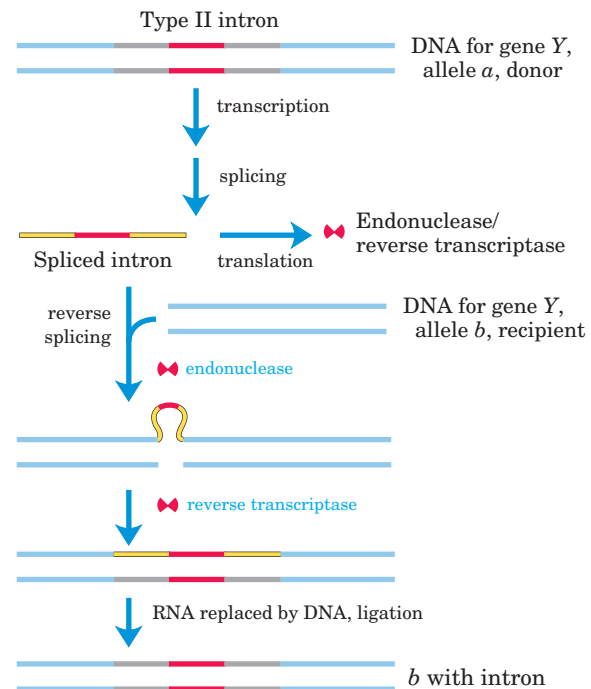
(a) Production of homing endonuclease



(b) Homing



(c) Retrohoming



Although the existence of this enzyme may not be surprising, the mechanism by which it acts is remarkable and unprecedented. Telomerase, like some other enzymes described in this chapter, contains both RNA and protein components. The RNA component is about 150 nucleotides long and contains about 1.5 copies of the appropriate C_yA_x telomere repeat. This region of the RNA acts as a template for synthesis of the T_xG_y strand of the telomere. Telomerase thereby acts as a cellular reverse transcriptase that provides the active site for RNA-dependent DNA synthesis. Unlike retroviral reverse transcriptases, telomerase copies only a small segment of RNA that it carries within itself. Telomere synthesis requires the 3' end of a chromosome as primer and proceeds in the usual $5' \rightarrow 3'$ direction. Having syn-

thesized one copy of the repeat, the enzyme repositions to resume extension of the telomere (Fig. 26–35a).

After extension of the T_xG_y strand by telomerase, the complementary C_yA_x strand is synthesized by cellular DNA polymerases, starting with an RNA primer (see Fig. 25–13). The single-stranded region is protected by specific binding proteins in many lower eukaryotes, especially those species with telomeres of less than a few hundred base pairs. In higher eukaryotes (including mammals) with telomeres many thousands of base pairs long, the single-stranded end is sequestered in a specialized structure called a **T loop**. The single-stranded end is folded back and paired with its complement in the double-stranded portion of the telomere. The formation of a T loop involves invasion of the 3' end

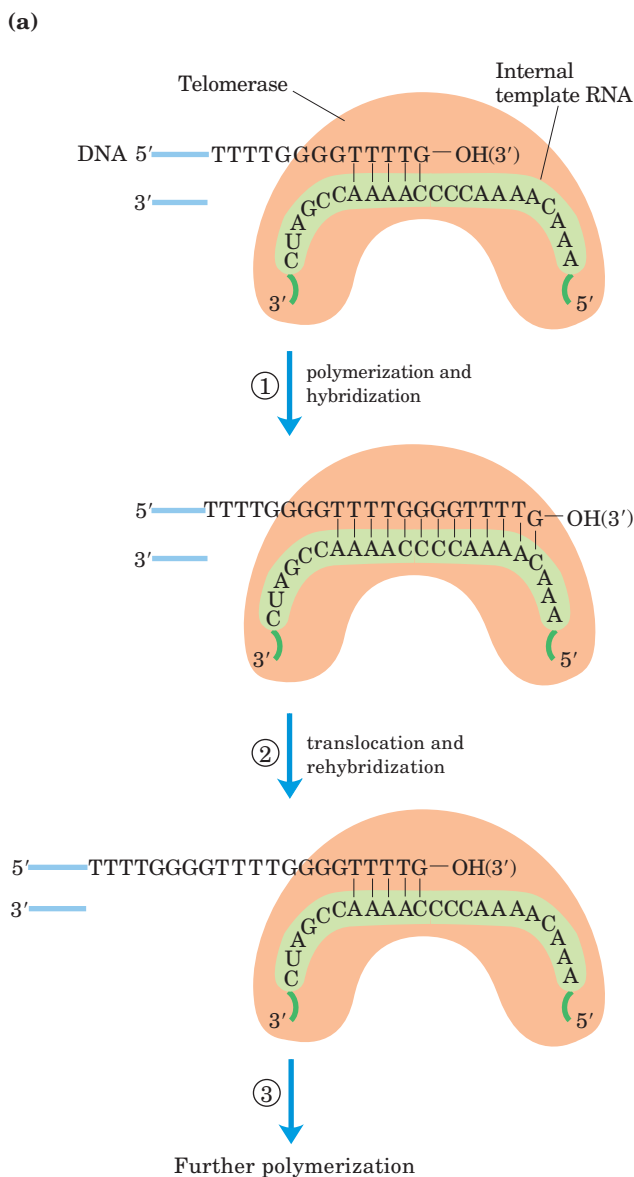
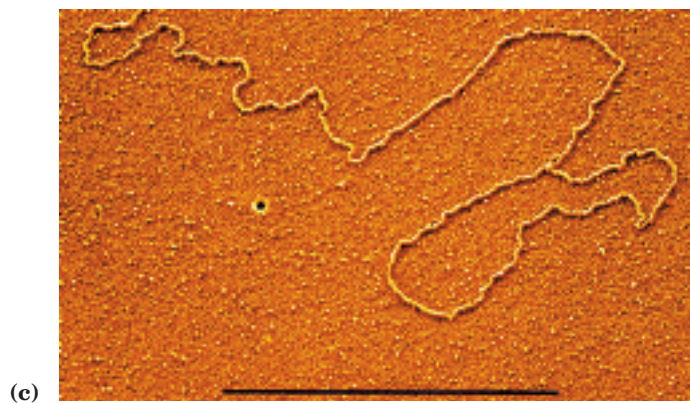
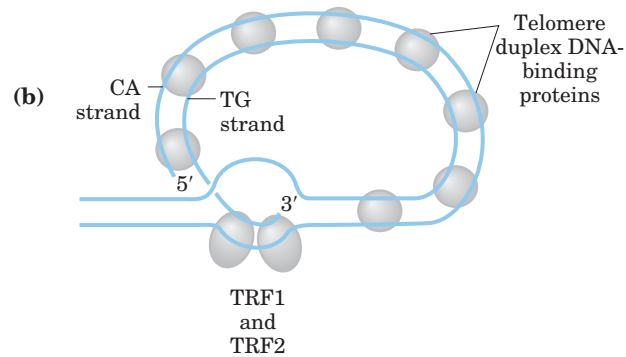


FIGURE 26–35 The TG strand and T loop of telomeres. The internal template RNA of telomerase binds to and base-pairs with the DNA's TG primer (T_xG_y). ① Telomerase adds more T and G residues to the TG primer, then ② repositions the internal template RNA to allow ③ the addition of more T and G residues. The complementary strand is synthesized by cellular DNA polymerases (not shown). (b) Proposed structure of T loops in telomeres. The single-stranded tail synthesized by telomerase is folded back and paired with its complement in the duplex portion of the telomere. The telomere is bound by several telomere-binding proteins, including TRF1 and TRF2 (telomere repeat binding factors). (c) Electron micrograph of a T loop at the end of a chromosome isolated from a mouse hepatocyte. The bar at the bottom of the micrograph represents a length of 5,000 bp.



of the telomere's single strand into the duplex DNA, perhaps by a mechanism similar to the initiation of homologous genetic recombination (see Fig. 25–31). In mammals, the looped DNA is bound by two proteins, TRF1 and TRF2, with the latter protein involved in formation of the T loop. T loops protect the 3' ends of chromosomes, making them inaccessible to nucleases and the enzymes that repair double-strand breaks (Fig. 26–35b).

In protozoans (such as *Tetrahymena*), loss of telomerase activity results in a gradual shortening of telomeres with each cell division, ultimately leading to the death of the cell line. A similar link between telomere length and cell senescence (cessation of cell division) has been observed in humans. In germ-line cells, which contain telomerase activity, telomere lengths are maintained; in somatic cells, which lack telomerase, they are not. There is a linear, inverse relationship between the length of telomeres in cultured fibroblasts and the age of the individual from whom the fibroblasts were taken: telomeres in human somatic cells gradually shorten as an individual ages. If the telomerase reverse transcriptase is introduced into human somatic cells in vitro, telomerase activity is restored and the cellular life span increases markedly.

Is the gradual shortening of telomeres a key to the aging process? Is our natural life span determined by the length of the telomeres we are born with? Further research in this area should yield some fascinating insights.

Some Viral RNAs Are Replicated by RNA-Dependent RNA Polymerase

Some *E. coli* bacteriophages, including f2, MS2, R17, and Q β , as well as some eukaryotic viruses (including influenza and Sindbis viruses, the latter associated with a form of encephalitis) have RNA genomes. The single-stranded RNA chromosomes of these viruses, which also function as mRNAs for the synthesis of viral proteins, are replicated in the host cell by an **RNA-dependent RNA polymerase (RNA replicase)**. All RNA viruses—with the exception of retroviruses—must encode a protein with RNA-dependent RNA polymerase activity because the host cells do not possess this enzyme.

The RNA replicase of most RNA bacteriophages has a molecular weight of ~210,000 and consists of four subunits. One subunit (M_r 65,000) is the product of the replicase gene encoded by the viral RNA and has the active site for replication. The other three subunits are host proteins normally involved in host-cell protein synthesis: the *E. coli* elongation factors Tu (M_r 30,000) and Ts (M_r 45,000) (which ferry amino acyl-tRNAs to the ribosomes) and the protein S1 (an integral part of the 30S ribosomal subunit).

These three host proteins may help the RNA replicase locate and bind to the 3' ends of the viral RNAs.

RNA replicase isolated from Q β -infected *E. coli* cells catalyzes the formation of an RNA complementary to the viral RNA, in a reaction equivalent to that catalyzed by DNA-dependent RNA polymerases. New RNA strand synthesis proceeds in the 5'→3' direction by a chemical mechanism identical to that used in all other nucleic acid synthetic reactions that require a template. RNA replicase requires RNA as its template and will not function with DNA. It lacks a separate proofreading endonuclease activity and has an error rate similar to that of RNA polymerase. Unlike the DNA and RNA polymerases, RNA replicases are specific for the RNA of their own virus; the RNAs of the host cell are generally not replicated. This explains how RNA viruses are preferentially replicated in the host cell, which contains many other types of RNA.

RNA Synthesis Offers Important Clues to Biochemical Evolution

The extraordinary complexity and order that distinguish living from inanimate systems are key manifestations of fundamental life processes. Maintaining the living state requires that *selected* chemical transformations occur very rapidly—especially those that use environmental energy sources and synthesize elaborate or specialized cellular macromolecules. Life depends on powerful and selective catalysts—enzymes—and on informational systems capable of both securely storing the blueprint for these enzymes and accurately reproducing the blueprint for generation after generation. Chromosomes encode the blueprint not for the cell but for the enzymes that construct and maintain the cell. The parallel demands for information and catalysis present a classic conundrum: what came first, the information needed to specify structure or the enzymes needed to maintain and transmit the information?

The unveiling of the structural and functional complexity of RNA led Carl Woese, Francis Crick, and Leslie Orgel to propose in the 1960s that this macromolecule might serve as both information carrier and catalyst. The discovery of catalytic RNAs took this proposal from



Carl Woese



Francis Crick



Leslie Orgel

conjecture to hypothesis and has led to widespread speculation that an “RNA world” might have been important in the transition from prebiotic chemistry to life (see Fig. 1–34). The parent of all life on this planet, in the sense that it could reproduce itself across the generations from the origin of life to the present, might have been a self-replicating RNA or a polymer with equivalent chemical characteristics.

How might a self-replicating polymer come to be? How might it maintain itself in an environment where the precursors for polymer synthesis are scarce? How could evolution progress from such a polymer to the modern DNA-protein world? These difficult questions can be addressed by careful experimentation, providing clues about how life on Earth began and evolved.

The probable origin of purine and pyrimidine bases is suggested by experiments designed to test hypotheses about prebiotic chemistry (pp. 32–33). Beginning with simple molecules thought to be present in the early atmosphere (CH_4 , NH_3 , H_2O , H_2), electrical discharges such as lightning generate, first, more reactive molecules such as HCN and aldehydes, then an array of amino acids and organic acids (see Fig. 1–33). When molecules such as HCN become abundant, purine and pyrimidine bases are synthesized in detectable amounts. Remarkably, a concentrated solution of ammonium cyanide, refluxed for a few days, generates adenine in yields of up to 0.5% (Fig. 26–36). Adenine may well have been the first and most abundant nucleotide constituent to appear on Earth. Intriguingly, most enzyme cofactors contain adenosine as part of their structure, although it plays no direct role in the cofactor function (see Fig. 8–41). This may suggest an evolutionary relationship, based on the simple synthesis of adenine from cyanide.

The RNA world hypothesis requires a nucleotide polymer to reproduce itself. Can a ribozyme bring about its own synthesis in a template-directed manner? The self-splicing rRNA intron of *Tetrahymena* (Fig. 26–26) catalyzes the reversible attack of a guanosine residue on the 5' splice junction (Fig. 26–37). If the 5' splice site and the internal guide sequence are removed from the intron, the rest of the intron can bind RNA strands paired with short oligonucleotides. Part of the remaining intact intron effectively acts as a template for the

alignment and ligation of the short oligonucleotides. The reaction is in essence a reversal of the attack of guanosine on the 5' splice junction, but the result is the synthesis of long RNA polymers from short ones, with the sequence of the product defined by an RNA template.

A self-replicating polymer would quickly use up available supplies of precursors provided by the relatively slow processes of prebiotic chemistry. Thus, from an early stage in evolution, metabolic pathways would be required to generate precursors efficiently, with the synthesis of precursors presumably catalyzed by ribozymes. The extant ribozymes found in nature have a limited repertoire of catalytic functions, and of the ribozymes that may once have existed, no trace is left. To explore the RNA world hypothesis more deeply, we need to know whether RNA has the potential to catalyze the many different reactions needed in a primitive system of metabolic pathways.

The search for RNAs with new catalytic functions has been aided by the development of a method that rapidly searches pools of random polymers of RNA and extracts those with particular activities: SELEX is nothing less than accelerated evolution in a test tube (Box 26–3). It has been used to generate RNA molecules that bind to amino acids, organic dyes, nucleotides, cyanocobalamin, and other molecules. Researchers have isolated ribozymes that catalyze ester and amide bond formation, $\text{S}_{\text{N}}2$ reactions, metallation of (addition of metal ions to) porphyrins, and carbon–carbon bond formation. The evolution of enzymatic cofactors with nucleotide “handles” that facilitate their binding to ribozymes might have further expanded the repertoire of chemical processes available to primitive metabolic systems.

As we shall see in the next chapter, some natural RNA molecules catalyze the formation of peptide bonds, offering an idea of how the RNA world might have been transformed by the greater catalytic potential of proteins. The synthesis of proteins would have been a major event in the evolution of the RNA world, but would also have hastened its demise. The information-carrying role of RNA may have passed to DNA because DNA is chemically more stable. RNA replicase and reverse transcriptase may be modern versions of enzymes that once played important roles in making the transition to the modern DNA-based system.

Molecular parasites may also have originated in an RNA world. With the appearance of the first inefficient self-replicators, transposition could have been a potentially important alternative to replication as a strategy for successful reproduction and survival. Early parasitic RNAs would simply hop into a self-replicating molecule via catalyzed transesterification, then passively undergo replication. Natural selection would have driven transposition to become site-specific, targeting sequences that did not interfere with the catalytic activities of the

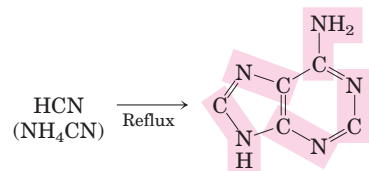


FIGURE 26–36 Possible prebiotic synthesis of adenine from ammonium cyanide. Adenine is derived from five molecules of cyanide, denoted by shading.

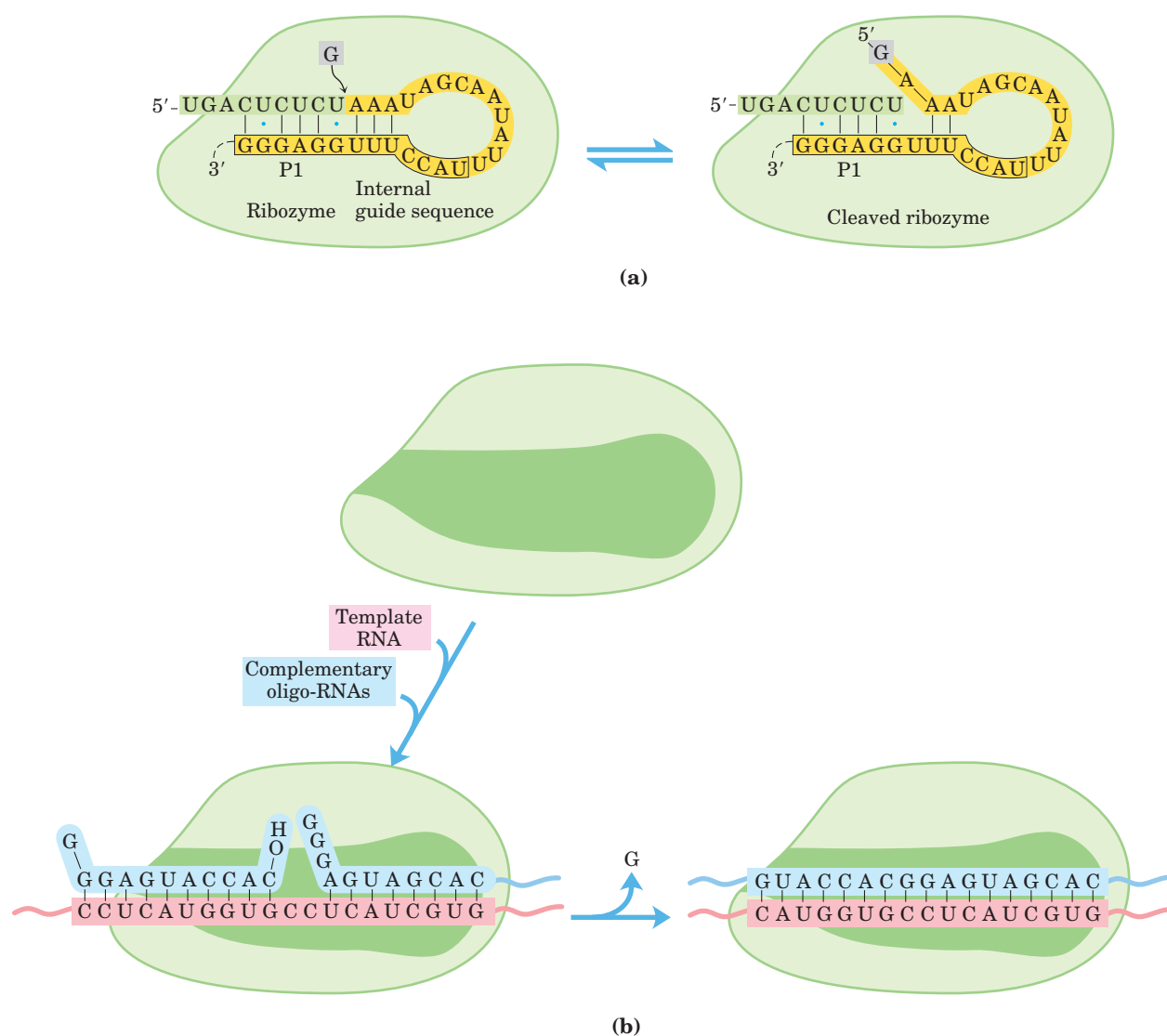


FIGURE 26-37 RNA-dependent synthesis of an RNA polymer from oligonucleotide precursors. **(a)** The first step in the removal of the self-splicing group I intron of the rRNA precursor of *Tetrahymena* is reversible attack of a guanosine residue on the 5' splice site. Only P1, the region of the ribozyme that includes the internal guide sequence (boxed) and the 5' splice site, is shown in detail; the rest of the ribozyme is represented as a green blob. The complete secondary structure of the ribozyme is shown in Figure 26-26. **(b)** If P1 is removed (shown as the darker green "hole"), the ribozyme retains both its three-

dimensional shape and its catalytic capacity. A new RNA molecule added *in vitro* can bind to the ribozyme in the same manner as does the internal guide sequence of P1 in **(a)**. This provides a template for further RNA polymerization reactions when oligonucleotides complementary to the added RNA base-pair with it. The ribozyme can link these oligonucleotides in a process equivalent to the *reversal* of the reaction in **(a)**. Although only one such reaction is shown in **(b)**, repeated binding and catalysis can result in the RNA-dependent synthesis of long RNA polymers.

host RNA. Replicators and RNA transposons could have existed in a primitive symbiotic relationship, each contributing to the evolution of the other. Modern introns, retroviruses, and transposons may all be vestiges of a "piggy-back" strategy pursued by early parasitic RNAs. These elements continue to make major contributions to the evolution of their hosts.

Although the RNA world remains a hypothesis, with many gaps yet to be explained, experimental evidence supports a growing list of its key elements. Further experimentation should increase our understanding. Important clues to the puzzle will be found in the workings of fundamental chemistry, in living cells, and perhaps on other planets.

BOX 26-3 WORKING IN BIOCHEMISTRY

The SELEX Method for Generating RNA Polymers with New Functions

SELEX (systematic evolution of ligands by exponential enrichment) is used to generate **aptamers**, oligonucleotides selected to tightly bind a specific molecular target. The process is generally automated to allow rapid identification of one or more aptamers with the desired binding specificity.

Figure 1 illustrates how SELEX is used to select an RNA species that binds tightly to ATP. In step ①, a random mixture of RNA polymers is subjected to “unnatural selection” by passing it through a resin to which ATP is attached. The practical limit for the complexity of an RNA mixture in SELEX is about 10^{15} different sequences, which allows for the complete randomization of 25 nucleotides ($4^{25} = 10^{15}$). When longer RNAs are used, the RNA pool used to initiate the search does not include all possible sequences. ② RNA polymers that pass through the column are discarded; ③ those that bind to ATP are washed from the column with salt solution and collected. ④ The collected RNA polymers are amplified by reverse transcriptase to make many DNA complements to the selected RNAs; then an RNA polymerase makes many RNA complements of the resulting DNA molecules. ⑤ This new pool of RNA is subjected to the same selection procedure, and the cycle is repeated a dozen or more times. At the end, only a few aptamers, in this

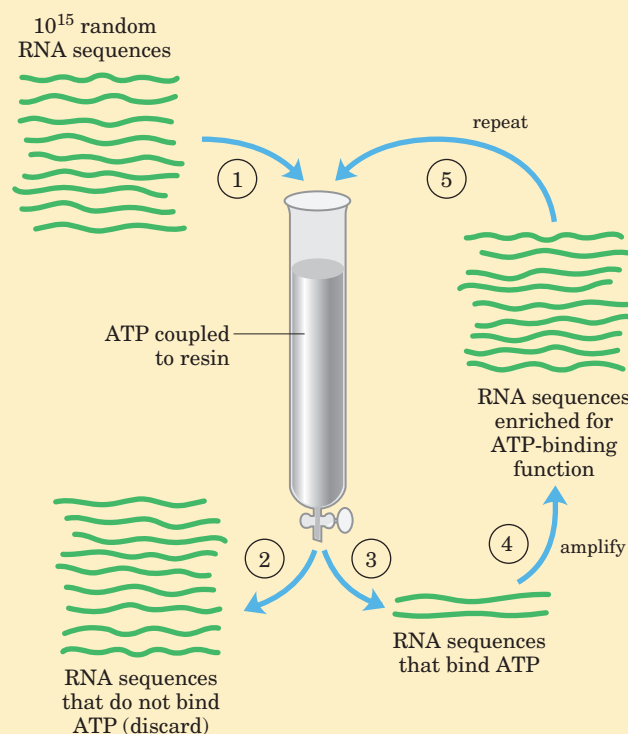


FIGURE 1 The SELEX procedure.

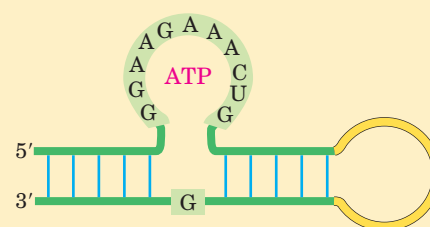


FIGURE 2 RNA aptamer that binds ATP. The shaded nucleotides are those required for the binding activity.

case RNA sequences with considerable affinity for ATP, remain.

Critical sequence features of an RNA aptamer that binds ATP are shown in Figure 2; molecules with this general structure bind ATP (and other adenosine nucleotides) with $K_d < 50 \mu\text{M}$. Figure 3 presents the three-dimensional structure of a 36 nucleotide RNA aptamer (shown as a complex with AMP) generated by SELEX. This RNA has the backbone structure shown in Figure 2.

In addition to its use in exploring the potential functionality of RNA, SELEX has an important practical side in identifying short RNAs with pharmaceutical uses. Finding an aptamer that binds specifically to every potential therapeutic target may be impossible, but the capacity of SELEX to rapidly select and amplify a specific oligonucleotide sequence from a highly complex pool of sequences makes this a promising approach for the generation of new therapies. For example, one could select an RNA that binds tightly to a receptor protein prominent in the plasma membrane of cells in a particular cancerous tumor. Blocking the activity of the receptor, or targeting a toxin to the tumor cells by attaching it to the aptamer, would kill the cells. SELEX also has been used to select DNA aptamers that detect anthrax spores. Many other promising applications are under development. ■

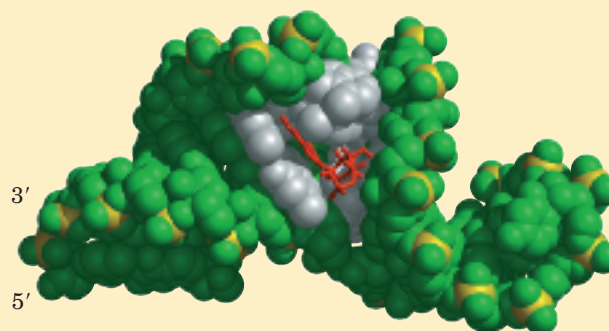


FIGURE 3 (Derived from PDB ID 1RAW.) RNA aptamer bound to AMP. The bases of the conserved nucleotides (forming the binding pocket) are white; the bound AMP is red.

SUMMARY 26.3 RNA-Dependent Synthesis of RNA and DNA

- RNA-dependent DNA polymerases, also called reverse transcriptases, were first discovered in retroviruses, which must convert their RNA genomes into double-stranded DNA as part of their life cycle. These enzymes transcribe the viral RNA into DNA, a process that can be used experimentally to form complementary DNA.
- Many eukaryotic transposons are related to retroviruses, and their mechanism of transposition includes an RNA intermediate.
- Telomerase, the enzyme that synthesizes the telomere ends of linear chromosomes, is a

specialized reverse transcriptase that contains an internal RNA template.

- RNA-dependent RNA polymerases, such as the replicases of RNA bacteriophages, are template-specific for the viral RNA.
- The existence of catalytic RNAs and pathways for the interconversion of RNA and DNA has led to speculation that an important stage in evolution was the appearance of an RNA (or an equivalent polymer) that could catalyze its own replication. The biochemical potential of RNAs can be explored by SELEX, a method for rapidly selecting RNA sequences with particular binding or catalytic properties.

Key Terms

Terms in bold are defined in the glossary.

transcription 995	repressor 1001	reverse transcriptase 1021
messenger RNA (mRNA) 995	footprinting 1002	retrovirus 1021
transfer RNA (tRNA) 995	transcription factors 1003	complementary DNA (cDNA) 1022
ribosomal RNA (rRNA) 995	ribozymes 1007	homing 1024
DNA-dependent RNA polymerase 996	primary transcript 1007	telomerase 1025
promoter 998	RNA splicing 1007	RNA-dependent RNA polymerase (RNA replicase) 1027
consensus sequence 998	5' cap 1008	aptamer 1030
cAMP receptor protein (CRP) 1001	spliceosome 1010	
	poly(A) tail 1011	

Further Reading

General

Jacob, F. & Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

A classic article that introduced many important ideas.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C.A., Krieger, M., Scott, M.P., Zipursky, S.L., & Darnell, J. (2003) *Molecular Cell Biology*, 5th edn, W. H. Freeman & Company, New York.

DNA-Directed RNA Synthesis

Conaway, J.W. & Conaway, R.C. (1999) Transcription elongation and human disease. *Annu. Rev. Biochem.* **68**, 301–320.

Conaway, J.W., Shilatifard, A., Dvir, A., & Conaway, R.C. (2000) Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* **25**, 375–380.

A particularly good summary of what is known about elongation factors.

DeHaseth, P.L., Zupancic, M.L., & Record, M.T., Jr. (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.* **180**, 3019–3025.

Friedberg, E.C. (1996) Relationships between DNA repair and transcription. *Annu. Rev. Biochem.* **65**, 15–42.

Kornberg, R.D. (1996) RNA polymerase II transcription control. *Trends Biochem. Sci.* **21**, 325–327.

Introduction to an issue of *Trends in Biochemical Sciences* that is devoted to RNA polymerase II.

Mooney, R.A., Artsimovitch, I., & Landick, R. (1998) Informational processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* **180**, 3265–3275.

Murakami, K.S. & Darst, S.A. (2003) Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13**, 31–39. This article and the two listed below explore the wealth of new structural information and what it tells us about RNA polymerase function.

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

Young, B.A., Gruber, T.M., & Gross, C.A. (2002) Views of transcription initiation. *Cell* **109**, 417–420.

RNA Processing

Beelman, C.A. & Parker, R. (1995) Degradation of mRNA in eukaryotes. *Cell* **81**, 179–183.

Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**, 333–360.

Chevalier, B.S. & Stoddard, B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acid Res.* **29**, 3757–3774.

Curcio, M.J. & Belfort, M. (1996) Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell* **84**, 9–12.

Frank, D.N. & Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA-processing ribozyme. *Annu. Rev. Biochem.* **67**, 153–180.

Jensen, T.H., Dower, K., Libri, D., & Rosbash, M. (2003) Early formation of mRNP: license for export or quality control? *Mol. Cell* **11**, 1129–1138.

A good summary of current ideas about the coupled processing and transport of eukaryotic mRNAs.

Kushner, S.R. (2002) mRNA decay in *Escherichia coli* comes of age. *J. Bacteriol.* **184**, 4658–4665.

Narlikar, G.J. & Herschlag, D. (1997) Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.* **66**, 19–59.

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

A description of current evidence for how processing is linked to the CTD of RNA polymerase II.

Sarkar, N. (1997) Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* **66**, 173–197.

Staley, J.P. & Guthrie, C. (1988) Mechanical devices of the spliceosome—motors, clocks, springs, and things. *Cell* **92**, 315–326.

RNA-Directed RNA or DNA Synthesis

Bishop, J.M. (1991) Molecular themes in oncogenesis. *Cell* **64**, 235–248.

A good overview of oncogenes; it introduces a series of more detailed reviews included in the same issue of *Cell*.

Blackburn, E.H. (1992) Telomerases. *Annu. Rev. Biochem.* **61**, 113–129.

Boeke, J.D. & Devine, S.E. (1998) Yeast retrotransposons: finding a nice, quiet neighborhood. *Cell* **93**, 1087–1089.

Collins, K. (1999) Ciliate telomerase biochemistry. *Annu. Rev. Biochem.* **68**, 187–218.

Frankel, A.D. & Young, J.A.T. (1998) HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1–25.

Greider, C.W. (1996) Telomere length regulation. *Annu. Rev. Biochem.* **65**, 337–365.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., & de Lange, T. (1999) Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514.

Lingner, J. & Cech, T.R. (1998) Telomerase and chromosome end maintenance. *Curr. Opin. Genet. Dev.* **8**, 226–232.

Temin, H.M. (1976) The DNA provirus hypothesis: the establishment and implications of RNA-directed DNA synthesis. *Science* **192**, 1075–1080.

Discussion of the original proposal for reverse transcription in retroviruses.

Zakian, V.A. (1995) Telomeres: beginning to understand the end. *Science* **270**, 1601–1607.

Ribozymes and Evolution

Bittker, J.A., Phillips, K.J., & Liu, D.R. (2002) Recent advances in the in vitro evolution of nucleic acids. *Curr. Opin. Chem. Biol.* **6**, 367–374.

DeRose, V.J. (2002) Two decades of RNA catalysis. *Chem. Biol.* **9**, 961–969.

Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., & Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325.

Review of progress toward the laboratory evolution of a self-replicating RNA.

Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature* **418**, 214–221.

Wilson, D.S. & Szostak, J.W. (1999) In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–648.

Yarus, M. (2002) Primordial genetics: phenotype of the ribocyte. *Annu. Rev. Genet.* **36**, 125–151.

Detailed speculations about what an RNA-based life form might have been like, and a good summary of the research behind the speculations.

Problems

1. RNA Polymerase (a) How long would it take for the *E. coli* RNA polymerase to synthesize the primary transcript for the *E. coli* genes encoding the enzymes for lactose metabolism (the 5,300 bp *lac* operon, considered in Chapter 28)? (b) How far along the DNA would the transcription “bubble” formed by RNA polymerase move in 10 seconds?

2. Error Correction by RNA Polymerases DNA polymerases are capable of editing and error correction, whereas the capacity for error correction in RNA polymerases appears to be quite limited. Given that a single base error in either replication or transcription can lead to an error in protein synthesis, suggest a possible biological explanation for this striking difference.

3. RNA Posttranscriptional Processing Predict the likely effects of a mutation in the sequence (5′)AAUAAA in a eukaryotic mRNA transcript.

4. Coding versus Template Strands The RNA genome of phage Q β is the nontemplate or coding strand, and when introduced into the cell it functions as an mRNA. Suppose the RNA replicase of phage Q β synthesized primarily template-strand RNA and uniquely incorporated this, rather than nontemplate strands, into the viral particles. What would be the fate of the template strands when they entered a new cell? What enzyme would such a template-strand virus need to include in the viral particles for successful invasion of a host cell?

5. The Chemistry of Nucleic Acid Biosynthesis Describe three properties common to the reactions catalyzed by DNA polymerase, RNA polymerase, reverse transcriptase, and RNA replicase. How is the enzyme polynucleotide phosphorylase similar to and different from these three enzymes?

6. RNA Splicing What is the minimum number of transesterification reactions needed to splice an intron from an mRNA transcript? Explain.

7. RNA Genomes The RNA viruses have relatively small genomes. For example, the single-stranded RNAs of retroviruses have about 10,000 nucleotides and the Q β RNA is only 4,220 nucleotides long. Given the properties of reverse transcriptase and RNA replicase described in this chapter, can you suggest a reason for the small size of these viral genomes?

8. Screening RNAs by SELEX The practical limit for the number of different RNA sequences that can be screened in a SELEX experiment is 10^{15} . (a) Suppose you are working with oligonucleotides 32 nucleotides in length. How many sequences exist in a randomized pool containing every sequence possible? (b) What percentage of these can be screened in a SELEX experiment? (c) Suppose you wish to select an RNA molecule that catalyzes the hydrolysis of a particular ester. From what you know about catalysis (Chapter 6), propose a SELEX strategy that might allow you to select the appropriate catalyst.

9. Slow Death The death cap mushroom, *Amanita phalloides*, contains several dangerous substances, including the lethal α -amanitin. This toxin blocks RNA elongation in consumers of the mushroom by binding to eukaryotic RNA polymerase II with very high affinity; it is deadly in concentrations as low as 10^{-8} M. The initial reaction to ingestion of the mushroom is gastrointestinal distress (caused by some of the other toxins). These symptoms disappear, but about 48 hours later, the mushroom-eater dies, usually from liver dysfunction. Speculate on why it takes this long for α -amanitin to kill.

10. Detection of Rifampicin-Resistant Strains of Tuberculosis Rifampicin is an important antibiotic used to treat tuberculosis, as well as other mycobacterial diseases.

Some strains of *Mycobacterium tuberculosis*, the causative agent of tuberculosis, are resistant to rifampicin. These strains become resistant through mutations that alter the *rpoB* gene, which encodes the β subunit of the RNA polymerase. Rifampicin cannot bind to the mutant RNA polymerase and so is unable to block the initiation of transcription. DNA sequences from a large number of rifampicin-resistant *M. tuberculosis* strains have been found to have mutations in a specific 69 bp region of *rpoB*. One well-characterized strain with rifampicin resistance has a single base pair alteration in *rpoB* that results in a single amino acid substitution in the β subunit: a His residue is replaced by an Asp residue.

(a) Based on your knowledge of protein chemistry (Chapters 3 and 4), suggest a technique that would allow detection of the rifampicin-resistant strain containing this particular mutant protein.

(b) Based on your knowledge of nucleic acid chemistry (Chapter 8), suggest a technique to identify the mutant form of *rpoB*.

Biochemistry on the Internet

11. The Ribonuclease Gene Human pancreatic ribonuclease has 128 amino acid residues.

(a) What is the minimum number of nucleotide pairs required to code for this protein?

(b) The mRNA expressed in human pancreatic cells was copied with reverse transcriptase to create a “library” of human DNA. The sequence of the mRNA coding for human pancreatic ribonuclease was determined by sequencing the complementary DNA (cDNA) from this library that included an open reading frame for the protein. Use the Entrez database system (www.ncbi.nlm.nih.gov/Entrez) to find the published sequence of this mRNA (search the nucleotide database for accession number D26129). What is the length of this mRNA?

(c) How can you account for the discrepancy between the size you calculated in (a) and the actual length of the mRNA?