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 the tRNAs of both bacteria and eukaryotes. Special-function RNAs are also processed.

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 .In a process called **RNA splicing**, the introns are removed from the primary transcript, and the exons are joined to form a continuous sequence that specifies a functional polypeptide. Eukaryotic mRNAs are also modified at each end. A modified residue called a 5′ cap 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 seem to be organized in association with each other and with the phosphorylated CTD of Pol II; each complex affects the function of the others. 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 comprising 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. The associated proteins modulate all aspects of the function and fate of the mRNA. These processes are outlined in **Figure 26-11.**

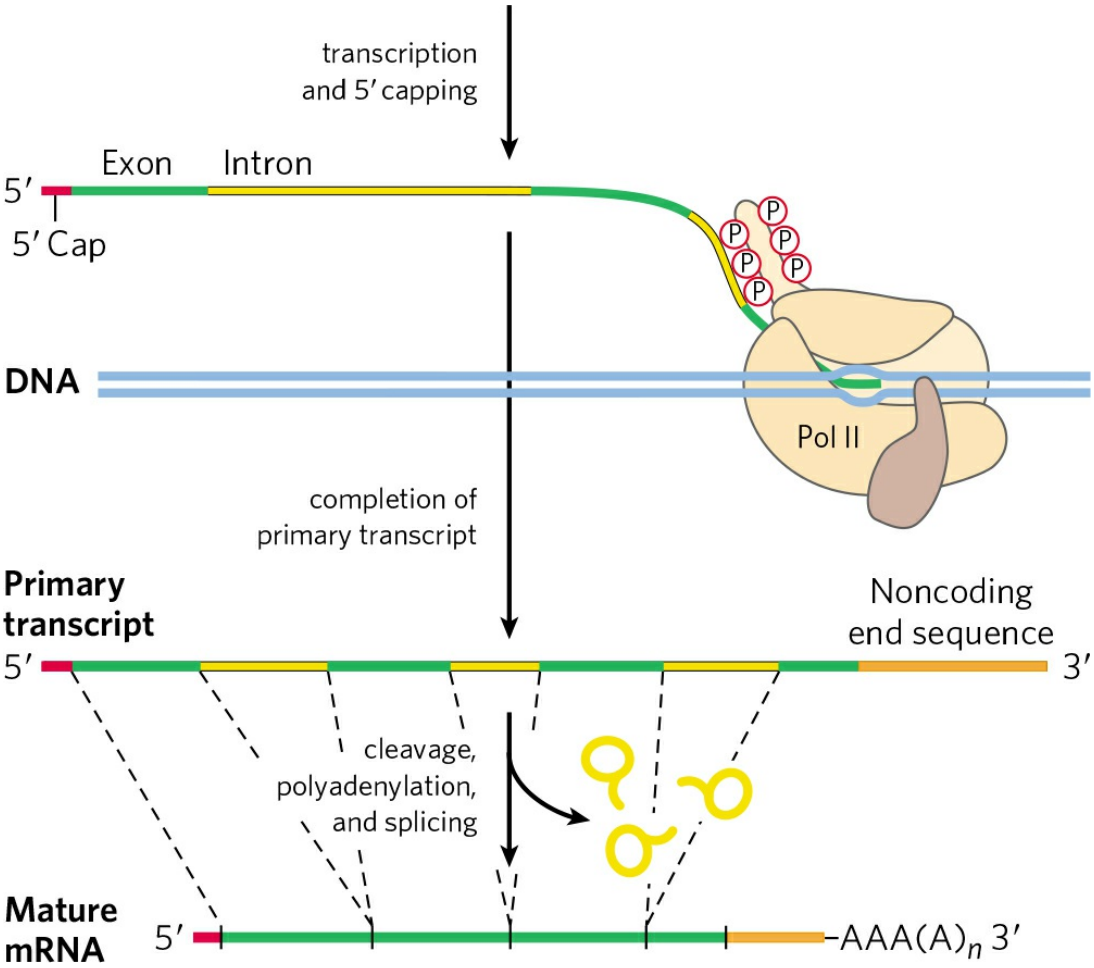


FIGURE 26-11 Formation of the primary transcript and its processing during maturation of mRNA in a eukaryotic cell. The 5′ cap (red) is added before synthesis of the primary transcript is complete. A noncoding end sequence (intron) following the last exon is shown in orange. Splicing can occur either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

The primary transcripts of bacterial 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. 26-22). Many of the special-function RNAs also undergo elaborate processing, often involving the removal of segments from one or both ends.

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

when its 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.

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. It also binds to a specific cap-binding complex of proteins and participates in binding of the mRNA to the ribosome to initiate translation

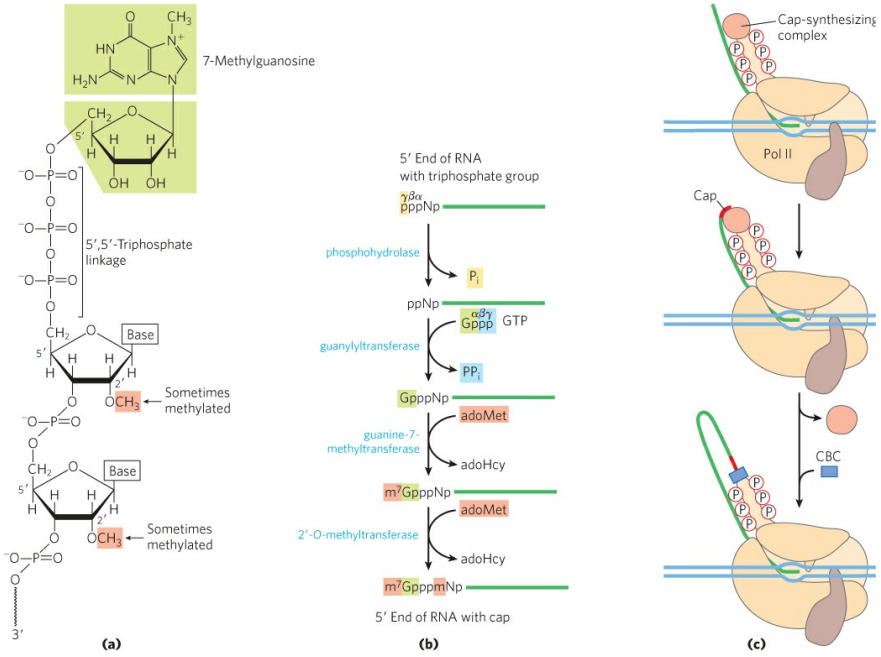
The 5′ cap is formed by condensation of a molecule of GTP with the triphosphate at the 5′ end of the transcript. The guanine is subsequently methylated at N-7, and additional methyl groups are often added at the 2′ hydroxyls of the first and second nucleotides adjacent to the cap (Fig. 26-12a). The methyl groups are derived from *S*-adenosylmethionine. All these reactions (Fig. 26-12b) occur very early in transcription, after the first 20 to 30 nucleotides of the transcript have been added. All four of the enzymes in the cap-synthesizing complex, 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 cap-synthesizing complex and bound by the cap-binding complex (Fig. 26-12c).

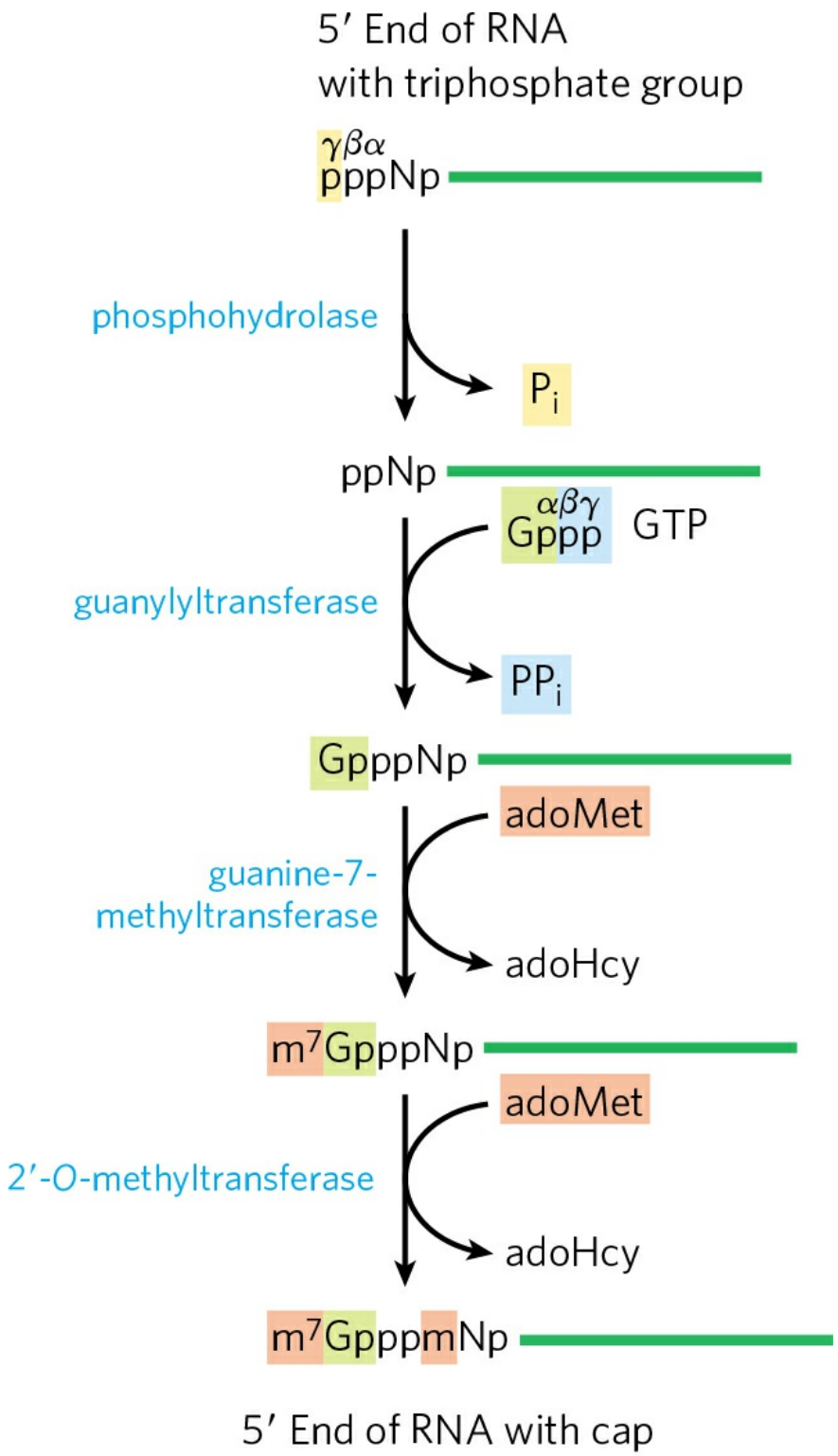
The 5′ cap does not provide complete protection of the transcript. The influenza virus has a genome consisting of eight segments of single-stranded RNA. Its genes are transcribed by a virally encoded RNA-dependent RNA polymerase, a heterotrimer consisting of subunits PA, PB1, and PB2. The virus needs no specialized enzymes for the synthesis of 5′ caps; instead, it borrows these structures from host-cell transcripts in a process termed “cap-snatching.” A capped host transcript is bound by the viral PB2 polymerase subunit and cleaved by an endonuclease in the PA subunit. The PB1 subunit uses the resulting capped oligonucleotide to prime viral RNA synthesis.

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 of the yeast *Saccharomyces cerevisiae* lack introns, but introns are more common in some other yeast species. Introns are also found in a few bacterial and archaeal 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 more than 700,000 nucleotides, with a median length of about 1,800. Genes of higher eukaryotes, including humans, typically have much more DNA devoted to introns than to exons. The ~20,000 genes of the human genome include more than 200,000 introns.




5′ End of RNA
with triphosphate group
γβα
pppNp
phosphohydrolase
Pi
ppNp
αβγ
Gppp
GTP
guanylyltransferase
PPi
GpppNp
adoMet
guanine-7-
methyltransferase
adoHcy
m7GpppNp
adoMet
2′-O-methyltransferase
adoHcy
m7GpppmNp
5′ End of RNA with cap

(b)

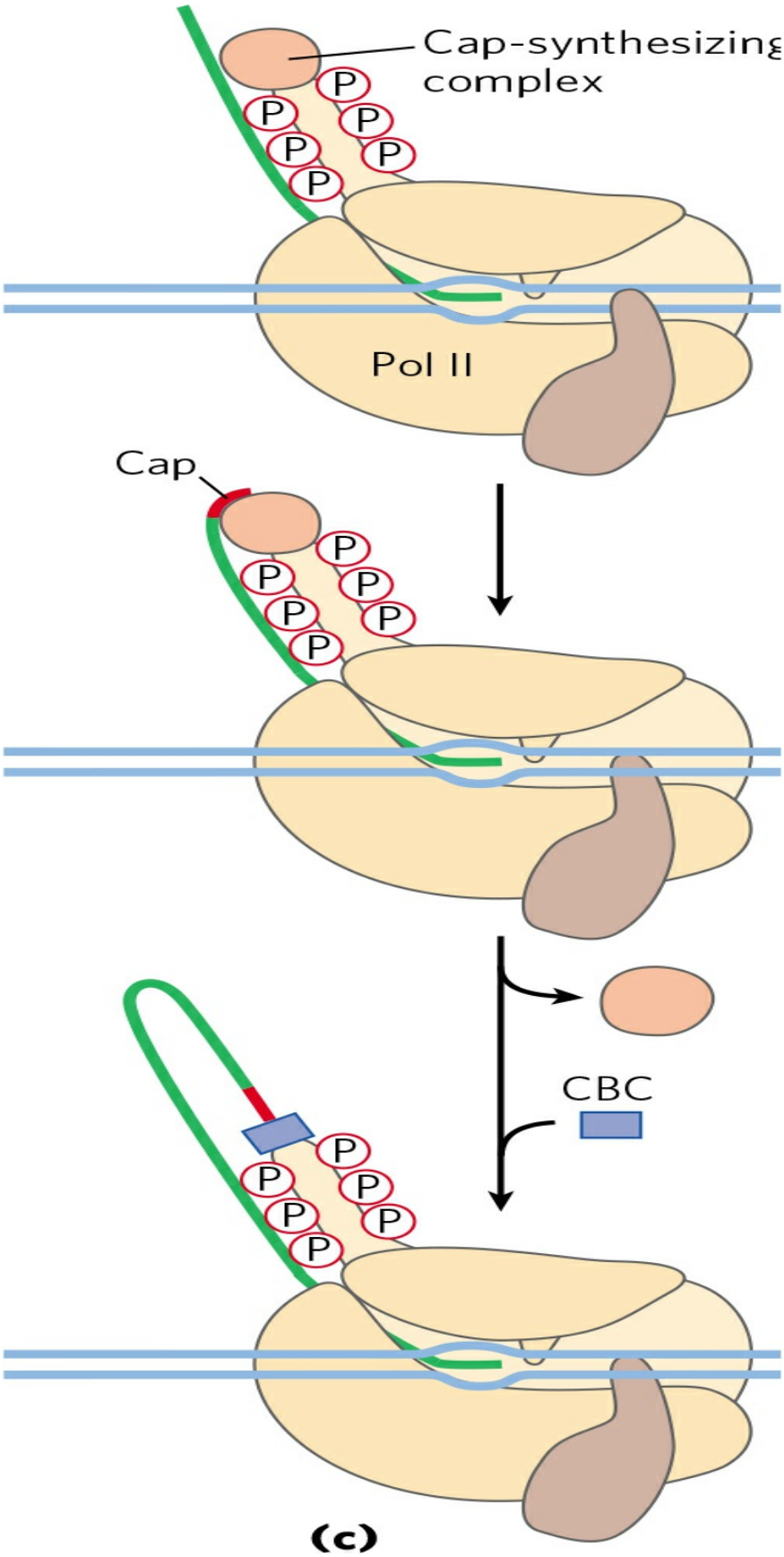

Cap-synthesizing
complex
P
Pol II
Cap
CBC
(c)

FIGURE 26-12 The 5′ cap of mRNA. (a) 7-Methylguanosine (m^7G) is joined to the 5′ end of almost all eukaryotic mRNAs in an unusual 5′,5′- triphosphate linkage. Methyl groups (light red) 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 requires four 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).

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 that code 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 rare 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)**, in which a ribose 2′- or 3′-hydroxyl group makes a nucleophilic attack on a phosphorus, and 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. 24-19) 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 and a major step forward in the understanding of how life probably evolved.

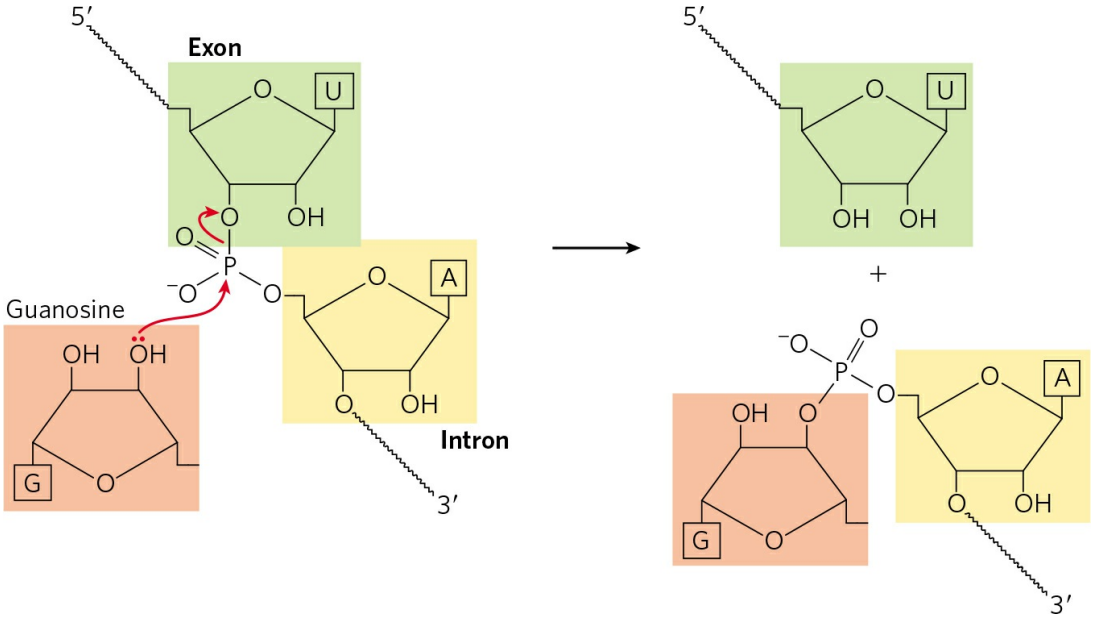


FIGURE 26-13 Transesterification reaction. Shown here is the first step in the two-step splicing of group I introns. In this example, the 3′ OH of a guanosine molecule acts as nucleophile, attacking the phosphodiester linkage between U and A residues at an exon-intron junction of an mRNA molecule (see Fig. 26-14).

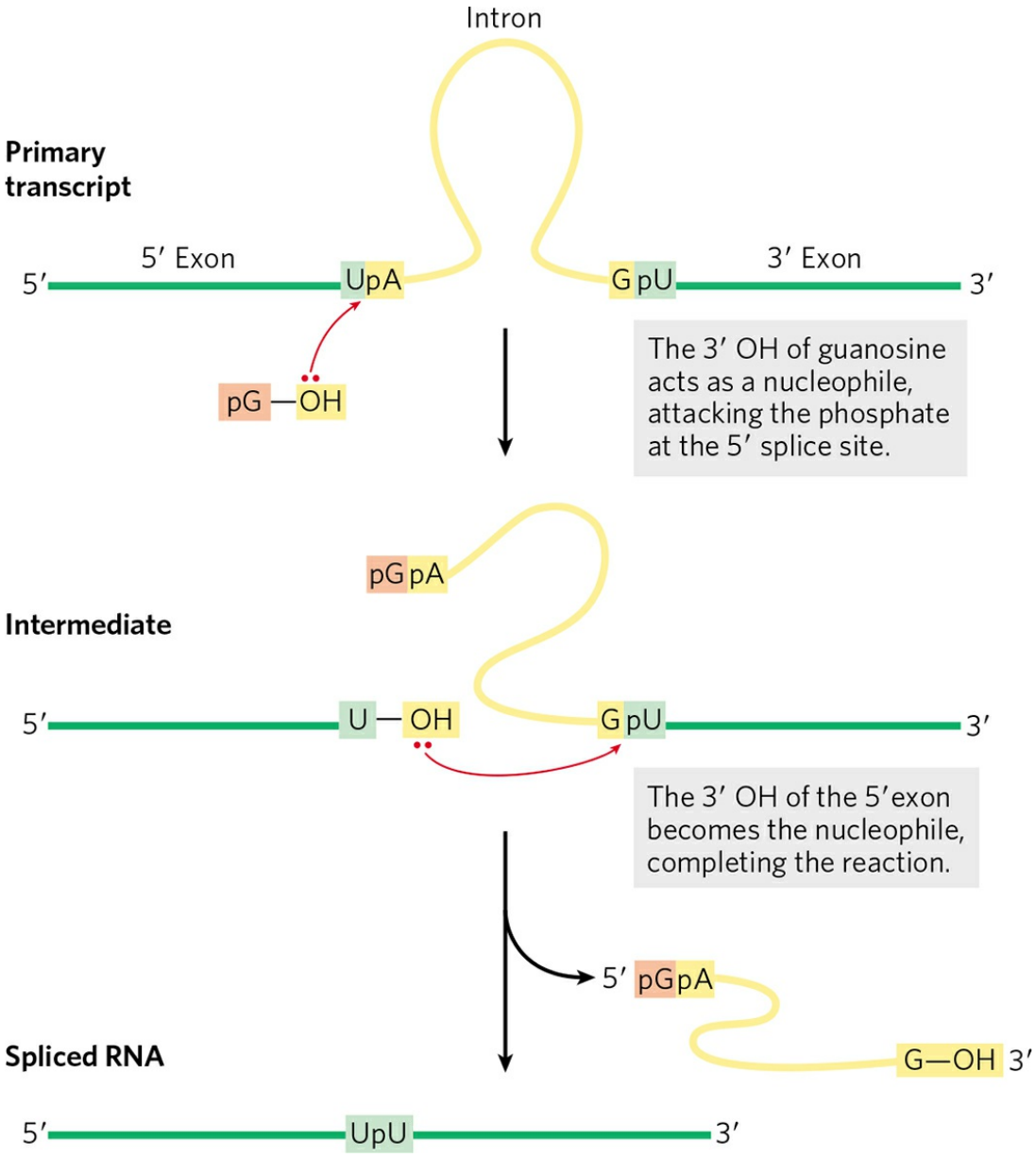


FIGURE 26-14 Splicing mechanism of group I introns. The nucleophile in the first step may be guanosine, GMP, GDP, or GTP. The spliced intron is eventually degraded.

In eukaryotes, most introns undergo splicing by the same lariat-forming mechanism as the group II introns. However, the intron splicing takes place within a large protein complex called a **spliceosome**, and these introns, the **spliceosomal introns**, are not assigned a group number. A spliceosome is made up of multiple specialized RNA-protein complexes called *s*mall *n*uclear *r*ibo*n*ucleo*p*roteins (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, U6) involved in splicing reactions are generally found in abundance in eukaryotic nuclei. In yeast, the various snRNPs include about 100 different proteins, most of which have close homologs in all other eukaryotes. In humans, these conserved protein components are augmented by more than 200 additional proteins. Spliceosomes are thus among the most complex macromolecular machines in any eukaryotic cell. The RNA components of a spliceosome are the catalysts of the various splicing steps. The overall complex can be considered a highly flexible nucleoprotein chaperone that can adapt to the great diversity in size and sequence of nuclear mRNAs.

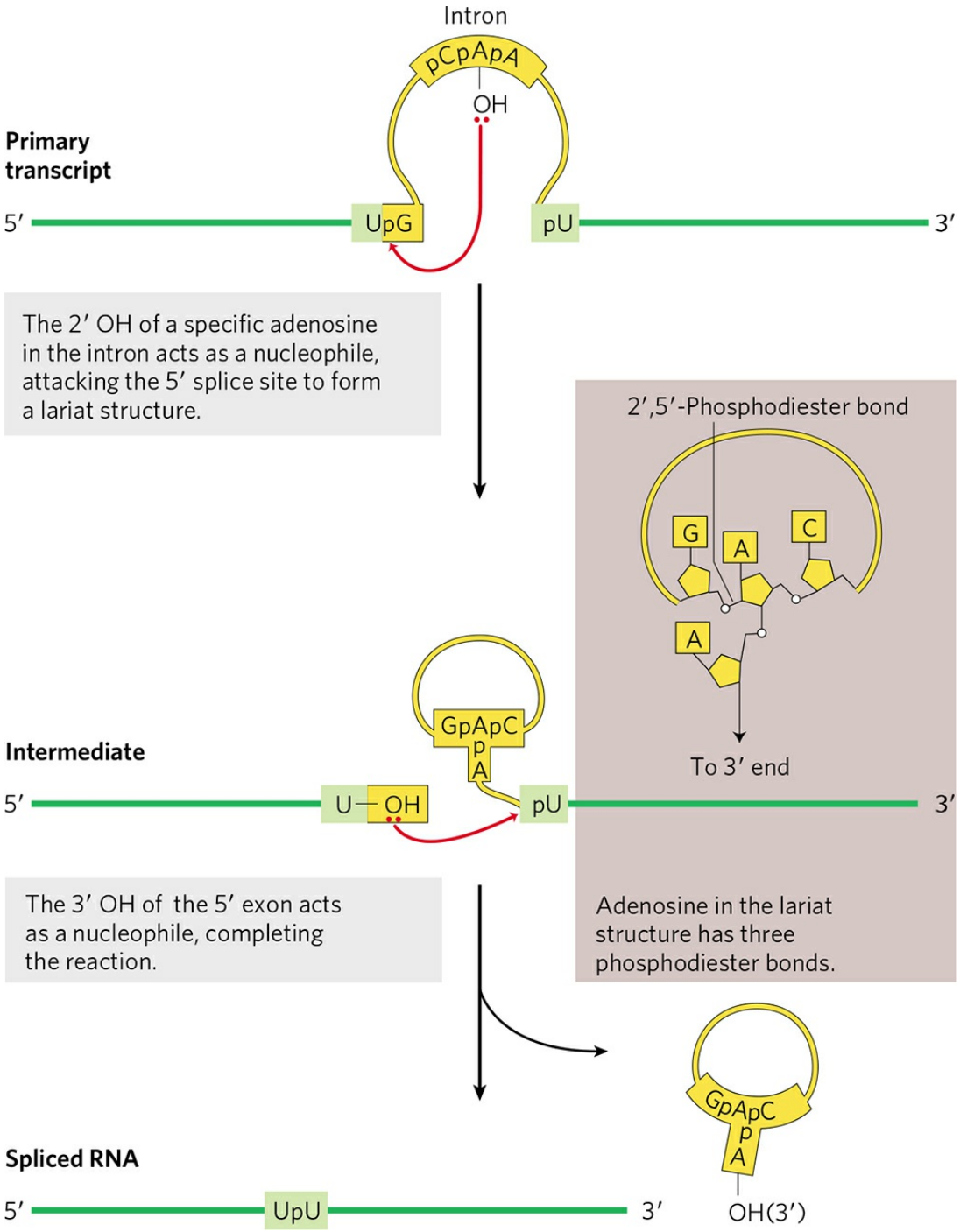


FIGURE 26-15 Splicing mechanism of group II introns. The chemistry is similar to that of group I intron splicing, except for the identity of the nucleophile in the first step and the formation of a lariatlike intermediate, in which one branch is a 2′,5′-phosphodiester bond.