O3 mRNA PROCESSING, hnRNPs AND snRNPs

Key Notes

Processing of mRNA

There is essentially no processing of prokaryotic mRNA; it can start to be translated before it has finished being transcribed. In eukaryotes, mRNA is synthesized by RNA Pol II as longer precursors (pre-mRNA), the population of different pre-mRNAs being called heterogeneous nuclear RNA (hnRNA). Specific proteins bind to hnRNA to form hnRNP and then small nuclear RNP (snRNP) particles interact with hnRNP to carry out some of the RNA processing events. Processing of eukaryotic hnRNA involves four events: 5′-capping, 3′-cleavage and polyadenylation, splicing and methylation.

hnRNP

RNA Pol II transcripts (hnRNA) complex with the three most abundant hnRNP proteins, the A, B and C proteins, to form hnRNP particles. These contain three copies of three tetramers and around 600–700 nucleotides of hnRNA. They assist RNA processing events.

snRNP particles

There are many uracil-rich snRNA molecules made by RNA Pol II which complex with specific proteins to form snRNPs. The most abundant are involved in splicing, and a large number define methylation sites in pre-rRNA. Those containing the sequence 5′-RA(U)$_n$GR-3′ bind eight common proteins in the cytoplasm, become hypermethylated and are imported back into the nucleus.

5′ Capping

This is the addition of a 7-methylguanosine nucleotide (m^7G) to the 5′-end of an RNA Pol II transcript when it is about 25 nt long. The m^7G, or cap, is added in reverse polarity (5′ to 5′), thus acting as a barrier to 5′-exonuclease attack, but it also promotes splicing, transport and translation.

3′ Cleavage and polyadenylation

Most eukaryotic pre-mRNAs are cleaved at a polyadenylation site and poly(A) polymerase (PAP) then adds a poly(A) tail of around 250 nt to generate the mature 3′-end.

Splicing

In eukaryotic pre-mRNA processing, intervening sequences (introns) that interrupt the coding regions (exons) are removed (spliced out), and the two flanking exons are joined. This splicing reaction occurs in the nucleus and requires the intron to have a 5′-GU, an AG-3′ and a branchpoint sequence. In a two-step reaction, the intron is removed as a tailed circular molecule, or lariat, and is degraded. Splicing involves the binding of snRNPs to the conserved sequences to form a spliceosome in which the cleavage and ligation reactions take place.

Pre-mRNA methylation

A small percentage of A residues in pre-mRNA, those in the sequence 5′-RRACX-3′, where R = purine, become methylated at the N6 position.

Related topics

RNA Pol II genes: promoters and enhancers (M4)

rRNA processing and ribosomes (O1)

Processing of mRNA

There appears to be little or no processing (see Topic O1) of mRNA transcripts in prokaryotes. In fact, ribosomes can assemble on, and begin to translate, mRNA molecules that have not yet been completely synthesized. Prokaryotic mRNA is degraded rapidly from the 5′-end and the first **cistron** (protein-coding region) can therefore only be translated for a limited amount of time. Some internal cistrons are partially protected by stem–loop structures that form at the 5′- and 3′-ends and provide a temporary barrier to exonucleases and can thus be translated more often before they are eventually degraded.

Because eukaryotic RNA Pol II transcribes such a wide variety of different genes, from the snRNA genes of 60–300 nt to the large *Antennapedia* gene, whose transcript can be over 100 kb in length, the collection of products made by this enzyme is referred to as **heterogeneous nuclear RNA (hnRNA)**. Those transcripts that will be processed to give mRNAs are called **pre-mRNAs**. Pre-mRNA molecules are processed to mature mRNA by 5′-capping, 3′-cleavage and polyadenylation, splicing and methylation.

hnRNP

The hnRNA synthesized by RNA Pol II is mainly pre-mRNA and rapidly becomes covered in proteins to form **heterogeneous nuclear ribonucleoprotein (hnRNP)**. The proteins involved have been classified as hnRNP proteins A–U. There are two forms of each of the three more abundant hnRNP proteins, the A, B and C proteins. Purification of this material from nuclei gives a fairly homogeneous preparation of 30–40S particles called hnRNP particles. These particles are about 20 nm in diameter and each contains about 600–700 nt of RNA complexed with three copies of three different tetramers. These tetramers are $(A_1)_3B_2$, $(A_2)_3B_1$ and $(C_1)_3C_2$. The hnRNP proteins are thought to help keep the hnRNA in a single-stranded form and to assist in the various RNA processing reactions.

snRNP particles

RNA Pol II also transcribes most **snRNAs** which complex with specific proteins to form **snRNPs**. These RNAs are rich in the base uracil and are thus denoted U1, U2, etc. The most abundant are those involved in **pre-mRNA splicing – U1, U2, U4, U5 and U6**. However, the list of snRNAs is growing, and the majority seem to be involved in determining the sites of methylation of pre-rRNA and are thus located in the nucleolus (see Topic O1). The major nucleoplasmic snRNPs are formed by the individual snRNAs complexing with a common set of eight proteins, which are small and basic, and a variable number of snRNP-specific proteins. These core proteins, known as the **Sm proteins** (after an antibody which recognizes them), require the sequence 5′-$RA(U)_nGR$-3′ to be present in a single-stranded region of the RNA. U6 does not have this sequence but it is usually base-paired to U4 which does. The snRNPs are formed as follows. They are synthesized in the nucleus by RNA Pol II and have a normal 5′-cap (see below). They are exported to the cytoplasm where they associate with the common core proteins and with other specific proteins. Their 5′-cap gains two methyl groups and they are then imported back into the nucleus where they function in splicing.

5′ Capping

Very soon after RNA Pol II starts making a transcript, and before the RNA chain is more than 20–30 nt long, the 5′-end is chemically modified by the addition of a **7-methylguanosine** (m^7G) residue (*Fig. 1*). This 5′ modification is called a **cap** and occurs by addition of a GMP nucleotide to the new RNA transcript in the reverse orientation compared with the normal 3′–5′ linkage, giving a

5′–5′ triphosphate bridge. The reaction is carried out by an enzyme called mRNA guanyltransferase and there can be subsequent methylations of the sugars on the first and second transcribed nucleotides, particularly in vertebrates. The cap structure forms a barrier to 5′-exonucleases and thus stabilizes the transcript, but the cap is also important in other reactions undergone by pre-mRNA and mRNA, such as splicing, nuclear transport and translation.

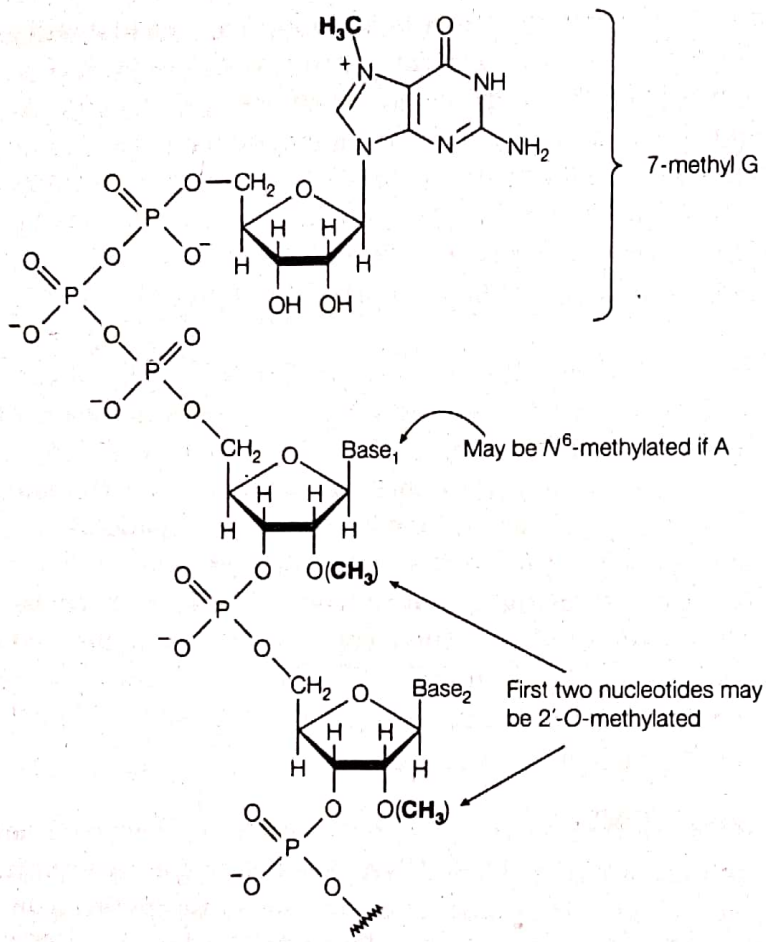

Fig. 1. The 5′ cap structure of eukaryotic mRNA.

3′ Cleavage and polyadenylation

In most pre-mRNAs, the mature 3′-end of the molecule is generated by cleavage followed by the addition of a run, or tail, of A residues which is called the **poly(A) tail**. This feature has allowed the purification of mRNA molecules from the other types of cellular RNAs, permitting the construction of cDNA libraries as described in Topics I1 and I2, from which specific genes have been isolated and their functions analyzed.

The cleavage and polyadenylation reaction requires that specific sequences be present in the DNA and its pre-mRNA transcript. These consist of a 5′-AAUAAA-3′, the **polyadenylation signal**, followed by a 5′-YA-3′, where Y = pyrimidine, in the next 11–20 nt (*Fig. 2a*). Downstream, a GU-rich sequence is often present. Collectively, these sequence elements make up the requirements of a **polyadenylation site**.

A number of specific protein factors recognize these sequence elements and bind to the pre-mRNA. When the complex has assembled, cleavage takes place and then one of the factors, **poly(A) polymerase (PAP)**, adds up to 250 A residues to the 3′-end of the cleaved pre-mRNA. The poly(A) tail on pre-mRNA is thought

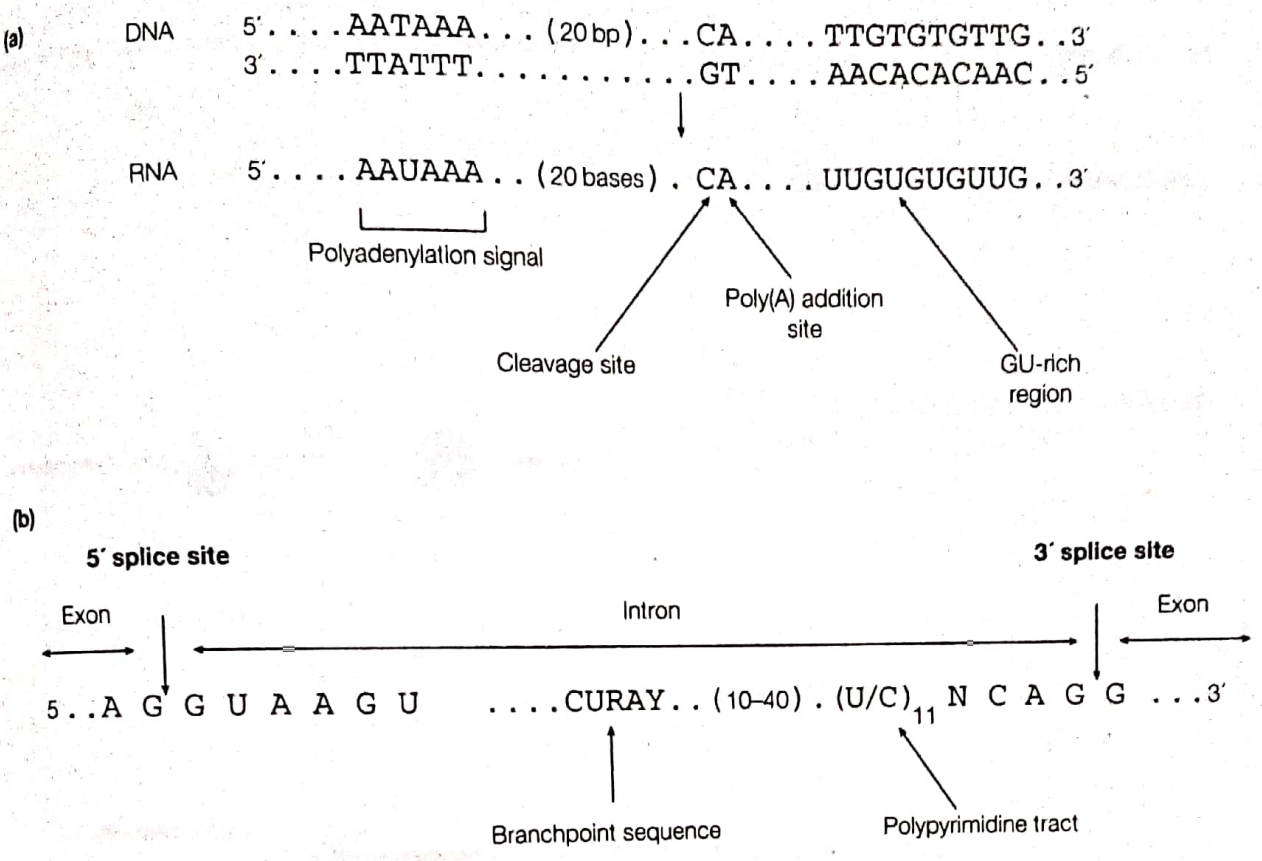

Fig. 2. Sequences of (a) a typical polyadenylation site and (b) the splice site consensus.

to help stabilize the molecule since a **poly(A)-binding protein** binds to it which should act to resist 3′-exonuclease action. In addition, the poly(A) tail may help in the translation of the mature mRNA in the cytoplasm. Histone pre-mRNAs do not get polyadenylated, but are cleaved at a special sequence to generate their mature 3′-ends.

Splicing

During the processing of pre-mRNA in eukaryotes, some sequences that are transcribed and which are upstream of the polyadenylation site are also eventually removed to create the mature mRNA. These sequences are cut out from central regions of the pre-mRNA and the outer portions joined. These intervening sequences, or **introns**, interrupt those sequences that will become adjacent regions in the mature mRNA, the **exons** which are usually the protein-coding regions of the mRNA. The process of cutting the pre-mRNA to remove the introns and joining together of the exons is called **splicing**. Like the polyadenylation process, it takes place in the nucleus before the mature mRNA can be exported to the cytoplasm.

Splicing also requires a set of specific sequences to be present (*Fig. 2b*). The 5′-end of almost all introns has the sequence 5′-GU-3′ and the 3′-end is usually 5′-AG-3′. The AG at the 3′-end is preceded by a pyrimidine-rich sequence called the **polypyrimidine tract**. About 10–40 residues upstream of the polypyrimidine tract is a sequence called the **branchpoint sequence** which is 5′-CURAY-3′ in vertebrates, where R = purine and Y = pyrimidine, but in yeast is the more specific sequence 5′-UACUAAC-3′.

Splicing has been shown to take place in a **two-step reaction** (*Fig. 3a*). First, the bond in front of the G at the 5′-end of the intron at the so-called **5′-splice** [illegible] by the 2′-hydroxyl group of the A residue of the branchpoint

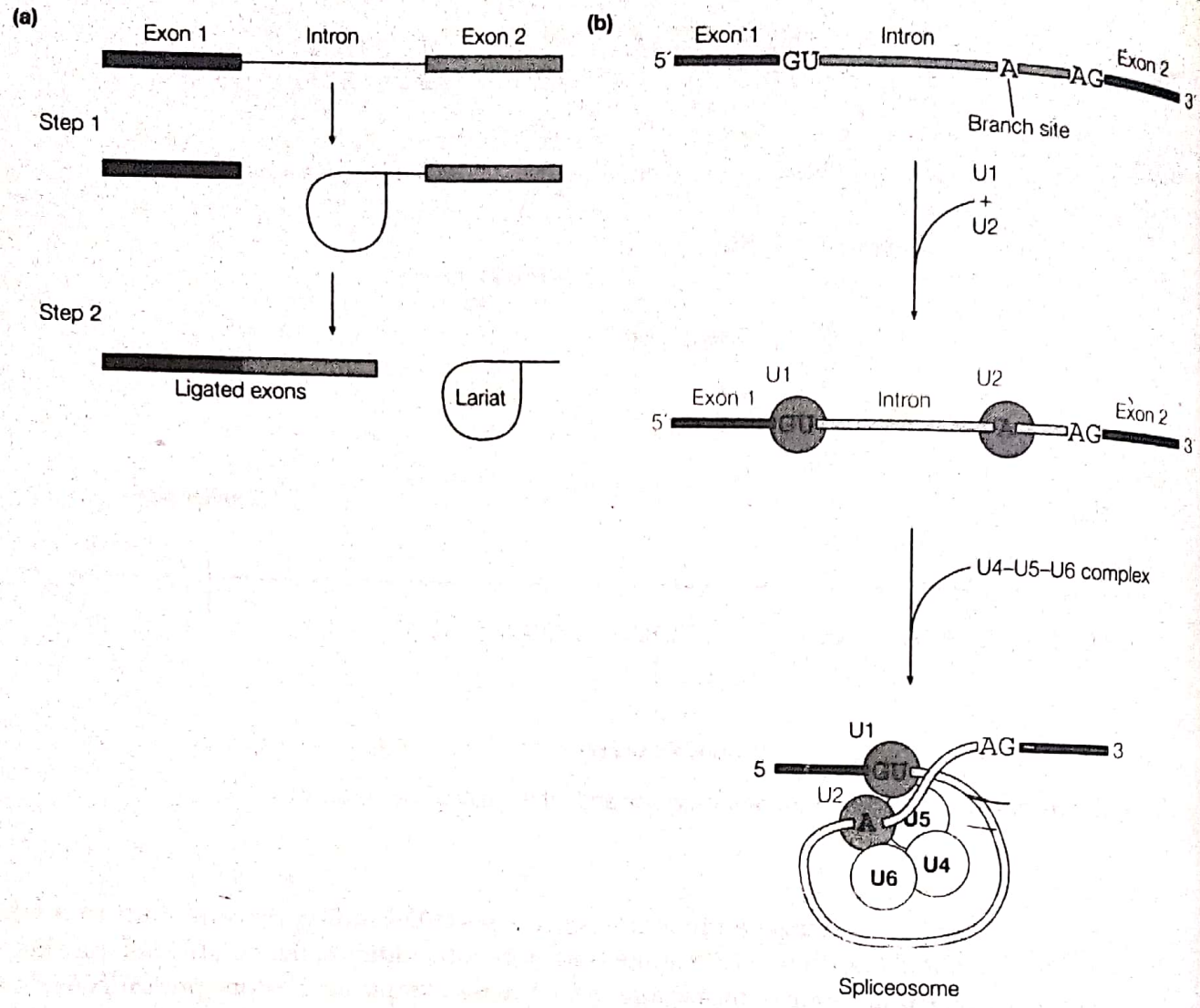

Fig. 3. Splicing of eukaryotic pre-mRNA. (a) The two-step reaction; (b) involvement of snRNPs in spliceosome formation.

sequence to create a tailed circular molecule called a **lariat** and free exon 1. In the second step, cleavage at the **3′-splice site** occurs after the G of the AG, as the two exon sequences are joined together. The intron is released in the lariat form and is eventually degraded.

The splicing process is catalyzed by the U1, U2, U4, U5 and U6 **snRNPs**, as well as other splicing factors. The RNA components of these snRNPs form base pairs with the various conserved sequences at the 5′- and 3′-splice sites and the branchpoint (*Fig. 3b*). Early in splicing, the 5′-end of the U1 snRNP binds to the 5′-splice site and then U2 binds to the branchpoint. The tri-snRNP complex of U4, U5 and U6 can then bind, and in so doing the intron is looped out and the 5′- and 3′-exons are brought into close proximity. The snRNPs interact with one another forming a complex which folds the pre-mRNA into the correct conformation for splicing. This complex of snRNPs and pre-mRNA which forms to hold the upstream and downstream exons close together while looping out the intron is called a **spliceosome**. After the spliceosome forms, a rearrangement takes place before the two-step splicing reaction can occur with release of the intron as a lariat.

Pre-mRNA methylation

The final modification or processing event that many pre-mRNAs undergo is specific methylation of certain bases. In vertebrates, the most common

methylation event is on the N6 position of A residues, particularly when these A residues occur in the sequence 5′-RRACX-3′, where X is rarely G. Up to 0.1% of pre-mRNA A residues are methylated, and the methylations seem to be largely conserved in the mature mRNA, though their function is unknown.

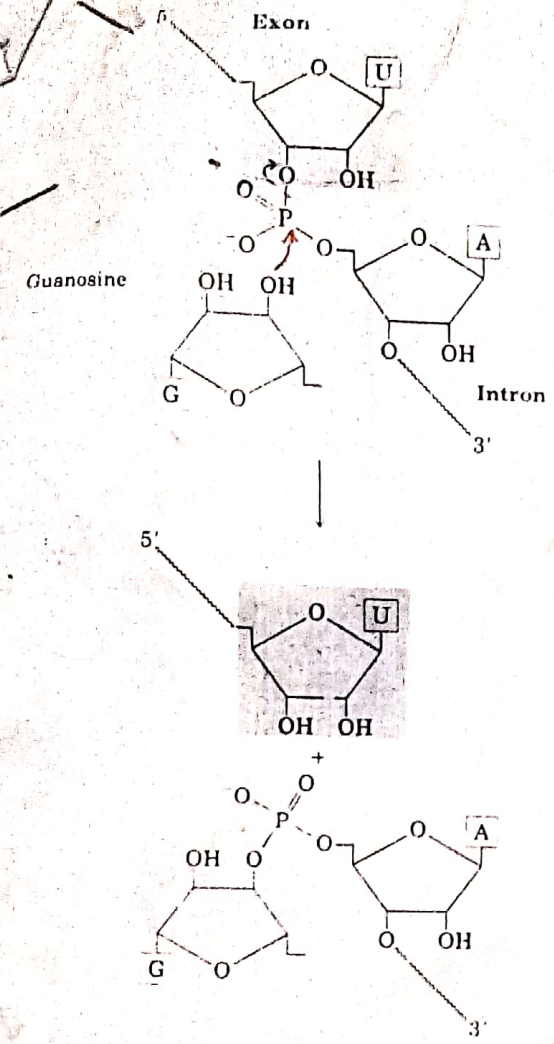

Figure 25–12 A transesterification reaction. This is the first step in the splicing of group I introns. Here, the 3′ OH of a guanosine molecule acts as nucleophile.

mechanisms involve two transesterification reaction steps (Fig. 25–12). A 2′- or 3′-hydroxyl group of a ribose makes a nucleophilic attack on a phosphorus, and in each step a new phosphodiester bond is formed at the expense of the old, maintaining an energy balance. Note that these reactions are very similar to the DNA breaking and rejoining reactions promoted by topoisomerases (Chapter 23) and site-specific recombinases (Chapter 24).

The group I splicing reaction requires a guanine nucleoside or nucleotide cofactor. This 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 forms a normal 3′,5′-phosphodiester bond with the 5′ end of the intron (Fig. 25–13). 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 pattern is similar except for the nucleophile in the first step. Instead of an external cofactor, the nucleophile is the 2′-hydroxyl group of an adenylate residue within the intron (Fig. 25–14). An unusual branched lariat structure is formed as an intermediate.

Attempts to identify the enzymes that promote splicing of group I and group II introns produced a major surprise; many of these introns are *self-splicing*—no protein enzymes are involved. This was first revealed in studies of the splicing mechanism of the group I rRNA intron from the ciliated protozoan *Tetrahymena thermophila* by Thomas Cech and colleagues in 1982. These workers proved that no proteins were involved by transcribing *Tetrahymena* DNA (including the intron) in

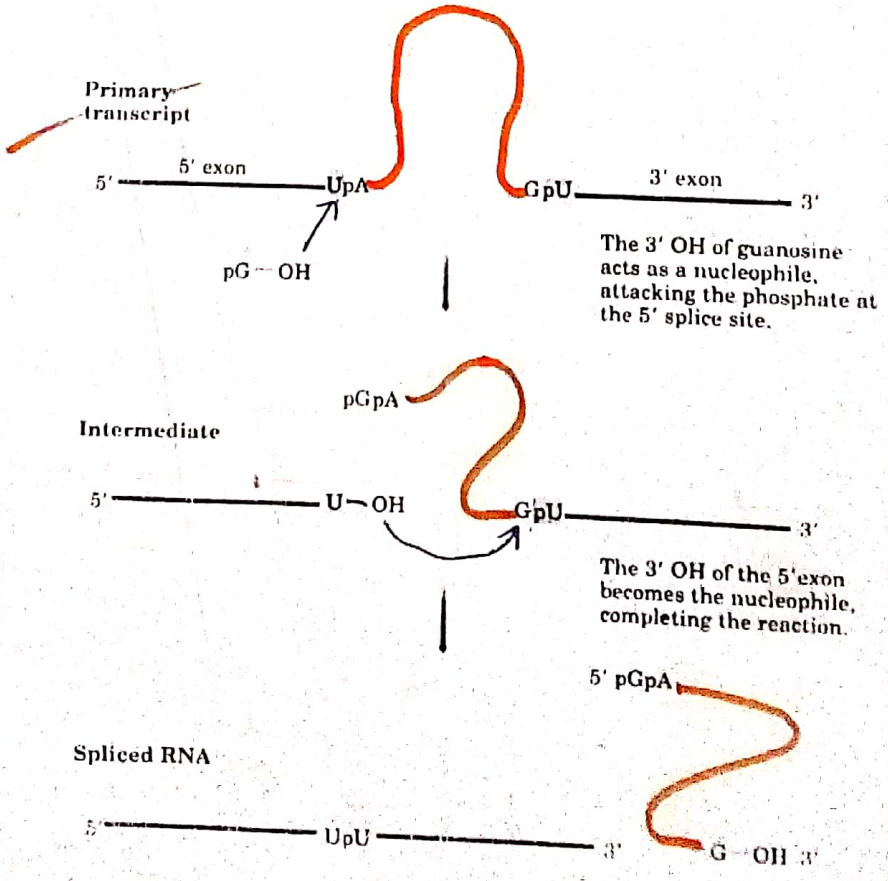

Figure 25–13 Splicing mechanism of group I introns. The nucleophile in the first step may be guanosine, GMP, GDP, or GTP.

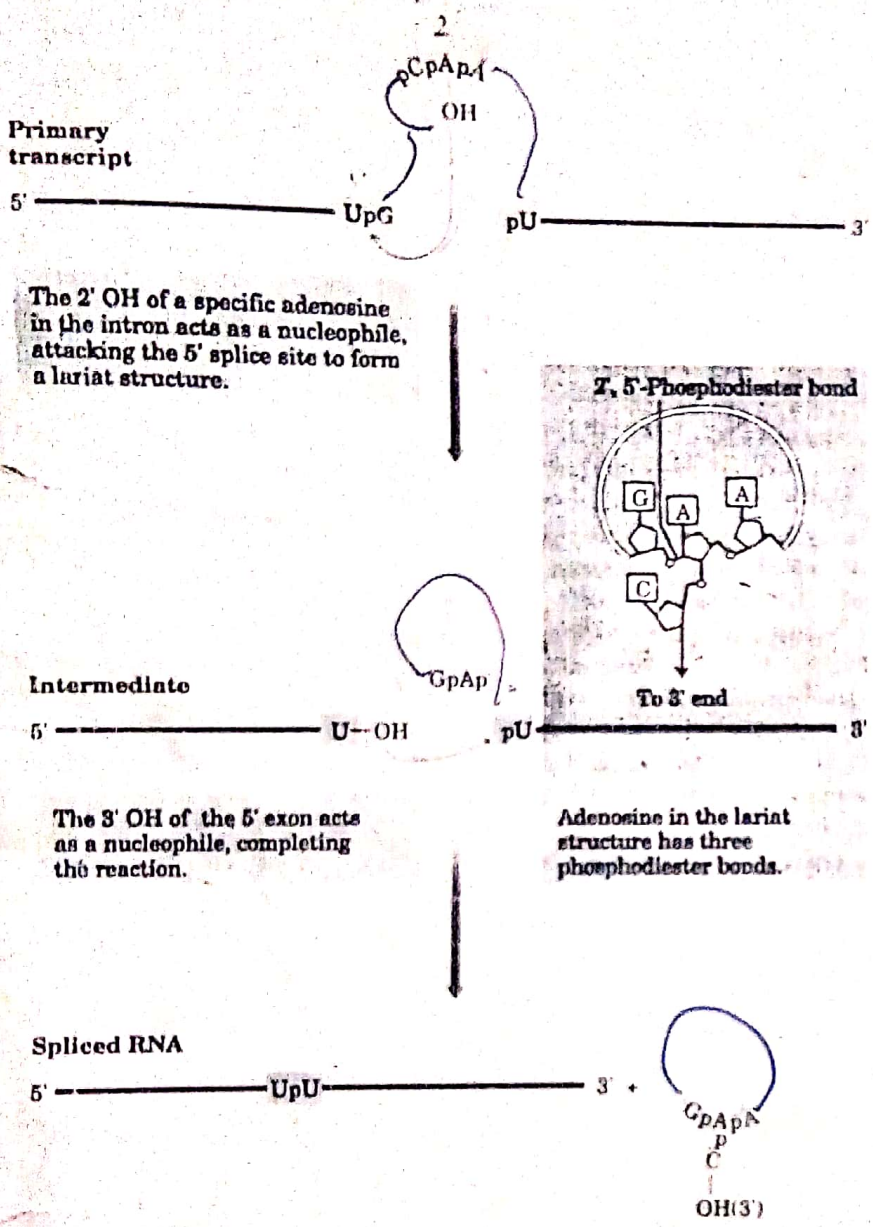

Figure 25–14 Splicing mechanism of group II introns. The chemistry is similar to that of group I intron splicing, except for the nucleophile in the first step and the novel lariatlike intermediate with one branch having a 2',5'-phosphodiester bond.

vitro using bacterial RNA polymerase. The resulting RNA spliced itself accurately even though it had never been in contact with any enzymes from *Tetrahymena*. The realization that RNAs, as well as proteins, could have catalytic functions was a milestone in thinking about biological systems. RNA catalysts are discussed in more detail later in this chapter.

The third and largest group of introns, found in nuclear mRNA primary transcripts, undergo splicing by the same lariat-formation mechanism as the group II introns. However, they are not self-splicing. Splicing requires the action of specialized RNA–protein complexes containing a class of eukaryotic RNAs called **small nuclear RNAs (snRNAs)**. Five snRNAs, U1, U2, U4, U5, and U6, are involved in splicing reactions. They are found in abundance in the nuclei of many eukaryotes, range in size from 106 (U6) to 189 (U2) nucleotides, and are complexed with proteins to form particles called *small nuclear ribonucleoproteins* (snRNPs, often referred to as "snurps"). The RNAs and proteins in snRNPs are highly conserved among vertebrates and invertebrates. Small nuclear RNAs similar to these are also found in yeast and molds.

FIGURE 11-12 Model for cleavage and polyadenylation of pre-mRNAs in mammalian cells. Cleavage-and-polyadenylation specificity factor (CPSF) binds to an upstream AAUAAA polyadenylation signal. CStF interacts with a downstream GU- or U-rich sequence and with bound CPSF, forming a loop in the RNA; binding of CFI and CFII help stabilize the complex. Binding of poly(A) polymerase (PAP) then stimulates cleavage at a poly(A) site, which usually is 10–35 nucleotides 3′ of the upstream polyadenylation signal. The cleavage factors are released, as is the downstream RNA cleavage product, which is rapidly degraded. Bound PAP then adds ≈12 A residues at a slow rate to the 3′-hydroxyl group generated by the cleavage reaction. Binding of poly(A)-binding protein II (PABII) to the initial short poly(A) tail accelerates the rate of addition by PAP. After 200–250 A residues have been added, PABII signals PAP to stop polymerization.