RNA SPLICING AND PROCESSING

5.24.1 Introduction

Key Terms

Pre-mRNA is used to describe the nuclear transcript that is processed by modification and splicing to give an mRNA.

RNA splicing is the process of excising the sequences in RNA that correspond to introns, so that the sequences corresponding to exons are connected into a continuous mRNA.

Heterogeneous nuclear RNA (hnRNA) comprises transcripts of nuclear genes made by RNA polymerase II; it has a wide size distribution and low stability.

An **hnRNP** is the ribonucleoprotein form of hnRNA (heterogeneous nuclear RNA), in which the hnRNA is complexed with proteins. Since pre-mRNAs are not exported until processing is complete, hnRNPs are found only in the nucleus.

Interrupted genes are found in all classes of organisms. They represent a minor proportion of the genes of the very lowest eukaryotes, but the vast majority of genes in higher eukaryotic genomes. Genes vary widely according to the numbers and lengths of introns, but a typical mammalian gene has 7-8 exons spread out over ~16 kb. The exons are relatively short (~100-200 bp), and the introns are relatively long (>1 kb) (see *Molecular Biology 1.2.7 Genes show a wide distribution of sizes*).

The discrepancy between the interrupted organization of the gene and the uninterrupted organization of its mRNA requires processing of the primary transcription product. The primary transcript has the same organization as the gene, and is sometimes called the **pre-mRNA**. Removal of the introns from pre-mRNA leaves a typical messenger of ~2.2 kb. The process by which the introns are removed is called **RNA splicing**.

Removal of introns is a major part of the production of RNA in all eukaryotes. (Although interrupted genes are relatively rare in lower eukaryotes such as yeast, the overall proportion underestimates the importance of introns, because most of the genes that are interrupted code for relatively abundant proteins. Splicing is therefore involved in the production of a greater proportion of total mRNA than would be apparent from analysis of the genome, perhaps as much as 50%.)

One of the first clues about the nature of the discrepancy in size between nuclear genes and their products in higher eukaryotes was provided by the properties of nuclear RNA. Its average size is much larger than mRNA, it is very unstable, and it has a much greater sequence complexity. Taking its name from its broad size distribution, it was called **heterogeneous nuclear RNA** (**hnRNA**). It includes pre-mRNA, but could also include other transcripts (that is, which are not ultimately processed to mRNA; for review see 17).

The physical form of hnRNA is a ribonucleoprotein particle (**hnRNP**), in which the

hnRNA is bound by proteins. As characterized *in vitro,* an hnRNP particle takes the form of beads connected by a fiber. The structure is summarized in **Figure 24.1**. The most abundant proteins in the particle are the core proteins, but other proteins are present at lower stoichiometry, making a total of ~20 proteins. The proteins typically are present at ~10^8 copies per nucleus, compared with ~10^6 molecules of hnRNA. Some of the proteins may have a structural role in packaging the hnRNA; several are known to shuttle between the nucleus and cytoplasm, and play roles in exporting the RNA or otherwise controlling its activity (for review see 249; 3428).

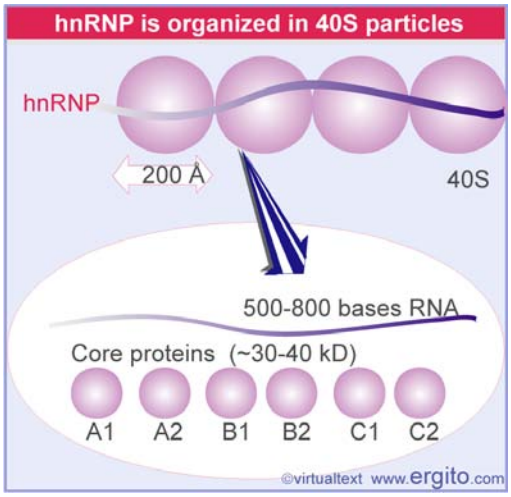

Figure 24.1 hnRNA exists as a ribonucleoprotein particle organized as a series of beads.

Splicing occurs in the nucleus, together with the other modifications that are made to newly synthesized RNAs. The process of expressing an interrupted gene is reviewed in **Figure 24.2**. The transcript is capped at the 5′ end (see *Molecular Biology 2.5.9 The 5′ end of eukaryotic mRNA is capped*), has the introns removed, and is polyadenylated at the 3′ end (see *Molecular Biology 2.5.10 The 3′ terminus is polyadenylated*). The RNA is then transported through nuclear pores to the cytoplasm, where it is available to be translated.

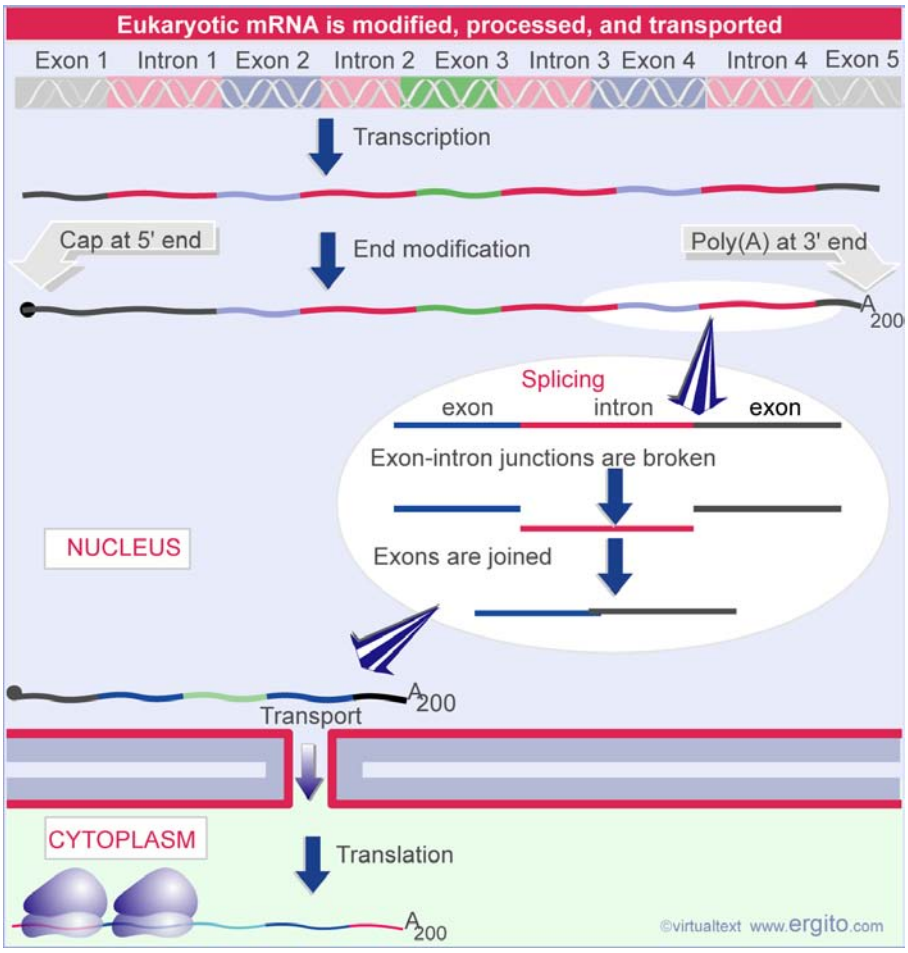

Figure 24.2 RNA is modified in the nucleus by additions to the 5′ and 3′ ends and by splicing to remove the introns. The splicing event requires breakage of the exon-intron junctions and joining of the ends of the exons. Mature mRNA is transported through nuclear pores to the cytoplasm, where it is translated.

With regard to the various processing reactions that occur in the nucleus, we should like to know at what point splicing occurs *vis*-À-*vis* the other modifications of RNA. Does splicing occur at a particular location in the nucleus; and is it connected with other events, for example, nucleocytoplasmic transport? Does the lack of splicing make an important difference in the expression of uninterrupted genes?

With regard to the splicing reaction itself, one of the main questions is how its specificity is controlled. What ensures that the ends of each intron are recognized in pairs so that the correct sequence is removed from the RNA? Are introns excised from a precursor in a particular order? Is the maturation of RNA used to *regulate* gene expression by discriminating among the available precursors or by changing the pattern of splicing?

We can identify several types of splicing systems:

- Introns are removed from the nuclear pre-mRNAs of higher eukaryotes by a

system that recognizes only short consensus sequences conserved at exon-intron boundaries and within the intron. This reaction requires a large splicing apparatus, which takes the form of an array of proteins and ribonucleoproteins that functions as a large particulate complex (the spliceosome). The mechanism of splicing involves transesterifications, and the catalytic center includes RNA as well as proteins.

- Certain RNAs have the ability to excise their introns autonomously. Introns of this type fall into two groups, as distinguished by secondary/tertiary structure. Both groups use transesterification reactions in which the RNA is the catalytic agent (see *Molecular Biology 5.26 Catalytic RNA*).

- The removal of introns from yeast nuclear tRNA precursors involves enzymatic activities that handle the substrate in a way resembling the tRNA processing enzymes, in which a critical feature is the conformation of the tRNA precursor. These splicing reactions are accomplished by enzymes that use cleavage and ligation.

Reviews

17. Lewin, B. (1975). *Units of transcription and translation: sequence components of hnRNA and mRNA.* Cell 4, 77-93.

249. Dreyfuss, G. et al. (1993). *hnRNP proteins and the biogenesis of mRNA.* Annu. Rev. Biochem. 62, 289-321.

3428. Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). *Messenger-RNA-binding proteins and the messages they carry.* Nat. Rev. Mol. Cell Biol. 3, 195-205.

RNA SPLICING AND PROCESSING

5.24.2 Nuclear splice junctions are short sequences

Key Terms

Splice sites are the sequences immediately surrounding the exon-intron boundaries.

The **GT-AG rule** describes the presence of these constant dinucleotides at the first two and last two positions of introns of nuclear genes.

Key Concepts

- Splice sites are the sequences immediately surrounding the exon-intron boundaries. They are named for their positions relative to the intron.
- The 5′ splice site at the 5′ (left) end of the intron includes the consensus sequence GU.
- The 3′ splice site at the 3′ (right) end of the intron includes the consensus sequence AG.
- The GU-AG rule (originally called the GT-AG rule in terms of DNA sequence) describes the requirement for these constant dinucleotides at the first two and last two positions of introns in pre-mRNAs.

To focus on the molecular events involved in nuclear intron splicing, we must consider the nature of the **splice sites**, the two exon-intron boundaries that include the sites of breakage and reunion.

By comparing the nucleotide sequence of mRNA with that of the structural gene, the junctions between exons and introns can be assigned. There is no extensive homology or complementarity between the two ends of an intron. However, the junctions have well conserved, though rather short, consensus sequences.

It is possible to assign a specific end to every intron by relying on the conservation of exon-intron junctions. They can all be aligned to conform to the consensus sequence given in **Figure 24.3**.

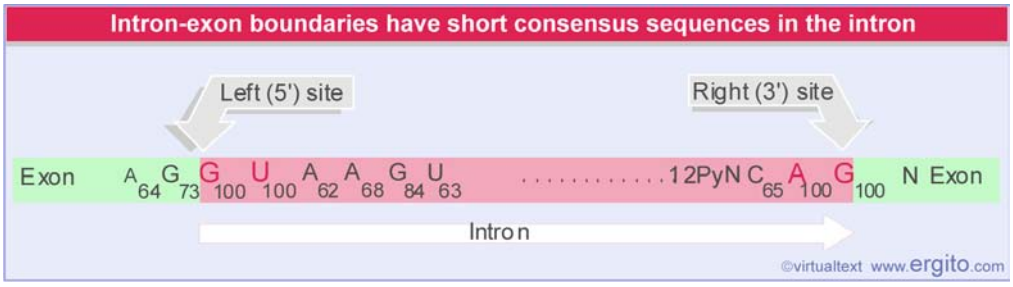

Figure 24.3 The ends of nuclear introns are defined by the GU-AG rule.

The subscripts indicate the percent occurrence of the specified base at each consensus position. High conservation is found only *immediately within the intron* at the presumed junctions. This identifies the sequence of a generic intron as:

GU##AG

Because the intron defined in this way starts with the dinucleotide GU and ends with the dinucleotide AG, the junctions are often described as conforming to the **GT-AG rule**. (This reflects the fact that the sequences were originally analyzed in terms of DNA, but of course the GT in the coding strand sequence of DNA becomes a GU in the RNA.)

Note that the two sites have different sequences and so they define the ends of the intron *directionally*. They are named proceeding from left to right along the intron as the 5′ splice site (sometimes called the left or donor site) and the 3′ splice site (also called the right or acceptor site). The consensus sequences are implicated as the sites recognized in splicing by point mutations that prevent splicing *in vivo* and *in vitro* (for review see 242; 243).

Reviews

242. Padgett, R. A. (1986). *Splicing of messenger RNA precursors*. Annu. Rev. Biochem. 55, 1119-1150.

243. Sharp, P. A. (1987). *Splicing of mRNA precursors*. Science 235, 766-771.

RNA SPLICING AND PROCESSING

5.24.3 Splice junctions are read in pairs

Key Concepts

- Splicing depends only on recognition of pairs of splice junctions.
- All 5′ splice sites are functionally equivalent, and all 3′ splice sites are functionally equivalent.

A typical mammalian mRNA has many introns. The basic problem of pre-mRNA splicing results from the simplicity of the splice sites, and is illustrated in **Figure 24.4**: what ensures that the correct pairs of sites are spliced together? The corresponding GU-AG pairs must be connected across great distances (some introns are >10 kb long). We can imagine two types of principle that might be responsible for pairing the appropriate 5′ and 3′ sites:

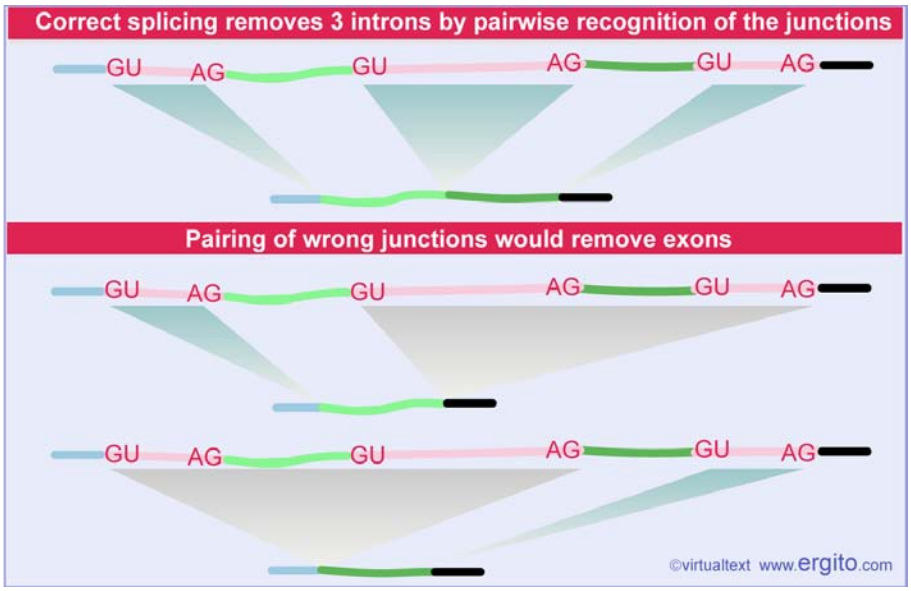

Figure 24.4 Splicing junctions are recognized only in the correct pairwise combinations.

- It could be an *intrinsic property* of the RNA to connect the sites at the ends of a particular intron. This would require matching of specific sequences or structures.
- Or all 5′ sites may be functionally equivalent and all 3′ sites may be similarly indistinguishable, but splicing could follow *rules* that ensure a 5′ site is always connected to the 3′ site that comes next in the RNA.

Neither the splice sites nor the surrounding regions have any sequence complementarity, which excludes models for complementary base pairing between

intron ends. And experiments using hybrid RNA precursors show that any 5′ splice site can in principle be connected to any 3′ splice site. For example, when the first exon of the early SV40 transcription unit is linked to the third exon of mouse β globin, the hybrid intron can be excised to generate a perfect connection between the SV40 exon and the β-globin exon. Indeed, this interchangeability is the basis for the exon-trapping technique described previously in **Figure 2.12**. Such experiments make two general points:

- *Splice sites are generic:* they do not have specificity for individual RNA precursors, and individual precursors do not convey specific information (such as secondary structure) that is needed for splicing.
- *The apparatus for splicing is not tissue specific;* an RNA can usually be properly spliced by any cell, whether or not it is usually synthesized in that cell. (We discuss exceptions in which there are tissue-specific alternative splicing patterns in *Molecular Biology 5.24.12 Alternative splicing involves differential use of splice junctions*.)

Here is a paradox. Probably all 5′ splice sites look similar to the splicing apparatus, and all 3′ splice sites look similar to it. *In principle any 5′ splice site may be able to react with any 3′ splice site.* But in the usual circumstances splicing occurs only between the 5′ and 3′ sites of the *same* intron. *What rules ensure that recognition of splice sites is restricted so that only the 5′ and 3′ sites of the same intron are spliced?*

Are introns removed in a specific *order* from a particular RNA? Using RNA blotting, we can identify nuclear RNAs that represent intermediates from which some introns have been removed. **Figure 24.5** shows a blot of the precursors to ovomucoid mRNA. There is a discrete series of bands, which suggests that splicing occurs via definite pathways. (If the seven introns were removed in an entirely random order, there would be more than 300 precursors with different combinations of introns, and we should not see discrete bands.)

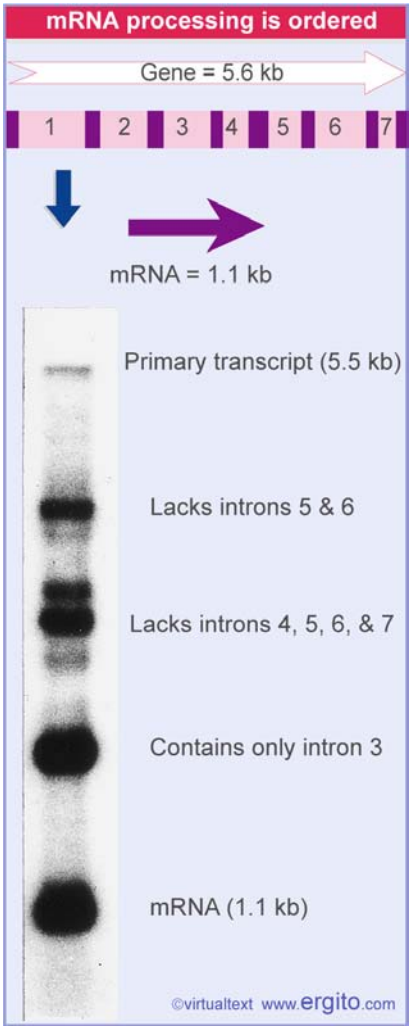

Figure 24.5 Northern blotting of nuclear RNA with an ovomucoid probe identifies discrete precursors to mRNA. The contents of the more prominent bands are indicated. Photograph kindly provided by Bert O#Malley.

There does not seem to be a *unique* pathway, since intermediates can be found in which different combinations of introns have been removed. However, there is evidence for a *preferred* pathway or pathways. When only one intron has been lost, it is virtually always 5 or 6. But either can be lost first. When two introns have been lost, 5 and 6 are again the most frequent, but there are other combinations. Intron 3 is never or very rarely lost at one of the first three splicing steps. From this pattern, we see that there is a preferred pathway in which introns are removed in the order 5/6, 7/4, 2/1, 3. But there are other pathways, since (for example), there are some molecules in which 4 or 7 is lost last. A caveat in interpreting these results is that we do not have proof that all these intermediates actually lead to mature mRNA.

The general conclusion suggested by this analysis is that the conformation of the RNA influences the accessibility of the splice sites. As particular introns are removed, the conformation changes, and new pairs of splice sites become available.

But the ability of the precursor to remove its introns in more than one order suggests that alternative conformations are available at each stage. Of course, the longer the molecule, the more structural options become available; and when we consider larger genes, it becomes difficult to see how specific secondary structures could control the reaction. One important conclusion of this analysis is that *the reaction does not proceed sequentially along the precursor*.

A simple model to control recognition of splice sites would be for the splicing apparatus to act in a processive manner. Having recognized a 5′ site, the apparatus might scan the RNA in the appropriate direction until it meets the next 3′ site. This would restrict splicing to adjacent sites. But this model is excluded by experiments that show that splicing can occur in *trans* as an intermolecular reaction under special circumstances (see *Molecular Biology 5.24.13 trans-splicing reactions use small RNAs*) or in RNA molecules in which part of the nucleotide chain is replaced by a chemical linker. This means that there cannot be a requirement for strict scanning along the RNA from the 5′ splice site to the 3′ splice site. Another problem with the scanning model is that it cannot explain the existence of alternative splicing patterns, where (for example) a common 5′ site is spliced to more than one 3′ site. The basis for proper recognition of correct splice site pairs remains incompletely defined.

RNA SPLICING AND PROCESSING

5.24.4 pre-mRNA splicing proceeds through a lariat

Key Terms

The **lariat** is an intermediate in RNA splicing in which a circular structure with a tail is created by a 5′-2′ bond.

The **branch site** is a short sequence just before the end of an intron at which the lariat intermediate is formed in splicing by joining the 5′ nucleotide of the intron to the 2′ position of an Adenosine.

A **transesterification** reaction breaks and makes chemical bonds in a coordinated transfer so that no energy is required.

Key Concepts

- Splicing requires the 5′ and 3′ splice sites and a branch site just upstream of the 3′ splice site.
- The branch sequence is conserved in yeast but less well conserved in higher eukaryotes.
- A lariat is formed when the intron is cleaved at the 5′ splice site, and the 5′ end is joined to a 2′ position at an A at the branch site in the intron.
- The intron is released as a lariat when it is cleaved at the 3′ splice site, and the left and right exons are then ligated together.
- The reactions occur by transesterifications in which a bond is transferred from one location to another.

The mechanism of splicing has been characterized *in vitro*, using systems in which introns can be removed from RNA precursors. Nuclear extracts can splice purified RNA precursors, which shows that the action of splicing is not linked to the process of transcription. Splicing can occur to RNAs that are neither capped nor polyadenylated. However, although the splicing reaction as such is independent of transcription or modification to the RNA, these events normally occur in a coordinated manner, and the efficiency of splicing may be influenced by other processing events.

The stages of splicing *in vitro* are illustrated in the pathway of **Figure 24.6**. We discuss the reaction in terms of the individual RNA species that can be identified, but remember that *in vivo* the species containing exons are not released as free molecules, but remain held together by the splicing apparatus (for review see 253).

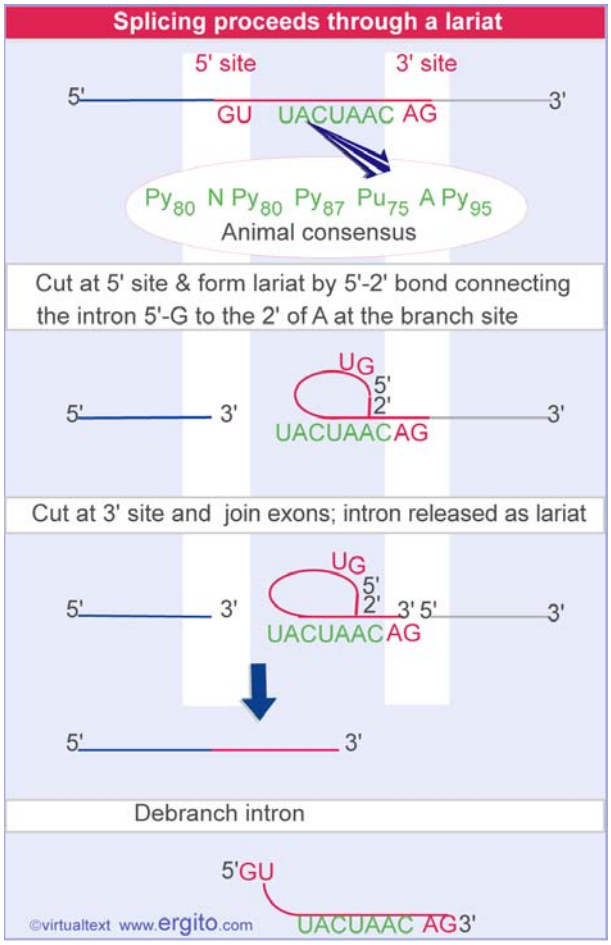

Figure 24.6 Splicing occurs in two stages. First the 5′ exon is cleaved off; then it is joined to the 3′ exon.

The first step is to make a cut at the 5′ splice site, separating the left exon and the right intron-exon molecule. The left exon takes the form of a linear molecule. The right intron-exon molecule forms a **lariat**, in which the 5′ terminus generated at the end of the intron becomes linked by a 5′–2′ bond to a base within the intron. The target base is an A in a sequence that is called the **branch site** (712).

Cutting at the 3′ splice site releases the free intron in lariat form, while the right exon is ligated (spliced) to the left exon. The cleavage and ligation reactions are shown separately in the figure for illustrative purposes, but actually occur as one coordinated transfer.

The lariat is then "debranched" to give a linear excised intron, which is rapidly degraded.

The sequences needed for splicing are the short consensus sequences at the 5′ and 3′ splice sites and at the branch site. Together with the knowledge that most of the sequence of an intron can be deleted without impeding splicing, this indicates that there is no demand for specific conformation in the intron (or exon).

The branch site plays an important role in identifying the 3′ splice site. The branch site in yeast is highly conserved, and has the consensus sequence UACUAAC. The branch site in higher eukaryotes is not well conserved, but has a preference for purines or pyrimidines at each position and retains the target A nucleotide (see **Figure 24.6**) (717).

The branch site lies 18-40 nucleotides upstream of the 3′ splice site. Mutations or deletions of the branch site in yeast prevent splicing. In higher eukaryotes, the relaxed constraints in its sequence result in the ability to use related sequences (called cryptic sites) when the authentic branch is deleted. Proximity to the 3′ splice site appears to be important, since the cryptic site is always close to the authentic site. A cryptic site is used only when the branch site has been inactivated. When a cryptic branch sequence is used in this manner, splicing otherwise appears to be normal; and the exons give the same products as wild type. *The role of the branch site therefore is to identify the nearest 3′ splice site as the target for connection to the 5′* splice site (713). This can be explained by the fact that an interaction occurs between protein complexes that bind to these two sites.

The bond that forms the lariat goes from the 5′ position of the invariant G that was at the 5′ end of the intron to the 2′ position of the invariant A in the branch site. This corresponds to the third A residue in the yeast UACUAAC box.

The chemical reactions proceed by **transesterification**: a bond is in effect *transferred* from one location to another. **Figure 24.7** shows that the first step is a nucleophilic attack by the 2′–OH of the invariant A of the UACUAAC sequence on the 5′ splice site. In the second step, the free 3′–OH of the exon that was released by the first reaction now attacks the bond at the 3′ splice site. Note that the number of phosphodiester bonds is conserved. There were originally two 5′–3′ bonds at the exon-intron splice sites; one has been replaced by the 5′–3′ bond between the exons, and the other has been replaced by the 5′–2′ bond that forms the lariat (for review see 251).

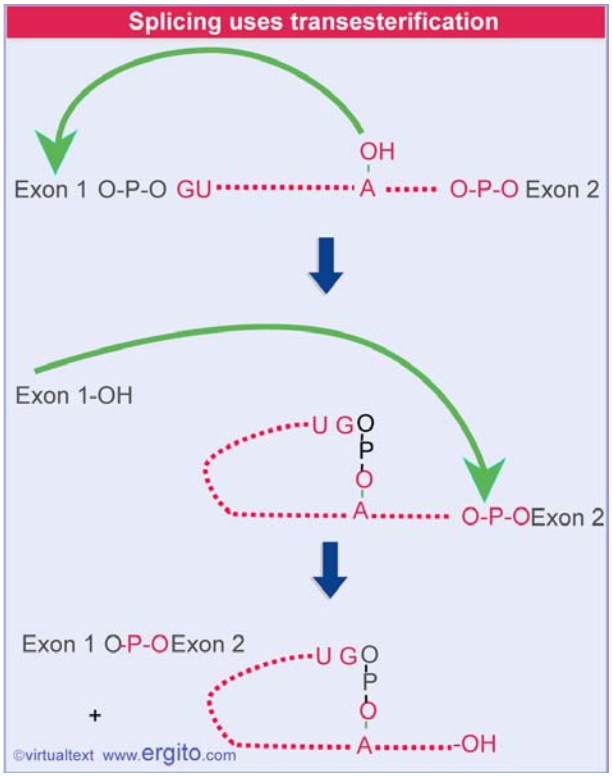

Figure 24.7 Nuclear splicing occurs by two transesterification reactions in which an OH group attacks a phosphodiester bond.

Reviews

251. Weiner, A. (1993). *mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism*. Cell 72, 161-164.

253. Sharp, P.A. (1994). *Split genes and RNA splicing*. Cell 77, 805-815.

References

712. Reed, R. and Maniatis, T. (1985). *Intron sequences involved in lariat formation during pre-mRNA splicing*. Cell 41, 95-105.

713. Reed, R. and Maniatis, T. (1986). *A role for exon sequences and splice-site proximity in splice-site selection*. Cell 46, 681-690.

717. Zhuang, Y. A., Goldstein, A. M., and Weiner, A. M. (1989). *UACUAAC is the preferred branch site for mammalian mRNA splicing*. Proc. Natl. Acad. Sci. USA 86, 2752-2756.

RNA SPLICING AND PROCESSING

5.24.5 snRNAs are required for splicing

Key Terms

A **small nuclear RNA (snRNA)** is one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions.

Small cytoplasmic RNAs (scRNA) are present in the cytoplasm and (sometimes are also found in the nucleus).

snRNPs (snurp) are small nuclear ribonucleoproteins (snRNAs associated with proteins).

scRNPs (scyrp) are small cytoplasmic ribonucleoproteins (scRNAs associated with proteins).

The **spliceosome** is a complex formed by the snRNPs that are required for splicing together with additional protein factors.

Anti-Sm is an autoimmune antiserum that defines the Sm epitope that is common to a group of proteins found in snRNPs that are involved in RNA splicing.

Key Concepts

- The five snRNPs involved in splicing are U1, U2, U5, U4, and U6.
- Together with some additional proteins, the snRNPs form the spliceosome.
- All the snRNPs except U6 contain a conserved sequence that binds the Sm proteins that are recognized by antibodies generated in autoimmune disease.

The 5′ and 3′ splice sites and the branch sequence are recognized by components of the splicing apparatus that assemble to form a large complex. This complex brings together the 5′ and 3′ splice sites before any reaction occurs, explaining why a deficiency in any one of the sites may prevent the reaction from initiating. The complex assembles sequentially on the pre-mRNA, and several intermediates can be recognized by fractionating complexes of different sizes. Splicing occurs only after all the components have assembled (719).

The splicing apparatus contains both proteins and RNAs (in addition to the pre-mRNA). The RNAs take the form of small molecules that exist as ribonucleoprotein particles. Both the nucleus and cytoplasm of eukaryotic cells contain many discrete small RNA species. They range in size from 100-300 bases in higher eukaryotes, and extend in length to ~1000 bases in yeast. They vary considerably in abundance, from 10^5-10^6 molecules per cell to concentrations too low to be detected directly.

Those restricted to the nucleus are called **small nuclear RNAs** (**snRNA**); those found in the cytoplasm are called **small cytoplasmic RNAs** (**scRNA**). In their natural state, they exist as ribonucleoprotein particles (snRNP and scRNP).

Colloquially, they are sometimes known as **snurps** and **scyrps**. There is also a class of small RNAs found in the nucleolus, called snoRNAs, which are involved in processing ribosomal RNA (see *Molecular Biology 5.24.22 Small RNAs are required for rRNA processing*).

The snRNPs involved in splicing, together with many additional proteins, form a large particulate complex, called the **spliceosome**. Isolated from the *in vitro* splicing systems, it comprises a 50-60S ribonucleoprotein particle. The spliceosome may be formed in stages as the snRNPs join, proceeding through several "presplicing complexes." The spliceosome is a large body, greater in mass than the ribosome.

Figure 24.8 summarizes the components of the spliceosome (3210). The 5 snRNAs account for more than a quarter of the mass; together with their 45 associated proteins, they account for almost half of the mass. Some 70 other proteins found in the spliceosome are described as splicing factors. They include proteins required for assembly of the spliceosome, proteins required for it to bind to the RNA substrate, and proteins involved in the catalytic process. In addition to these proteins, another ~30 proteins associated with the spliceosome have been implicated in acting at other stages of gene expression, suggesting that the spliceosome may serve as a coordinating apparatus.

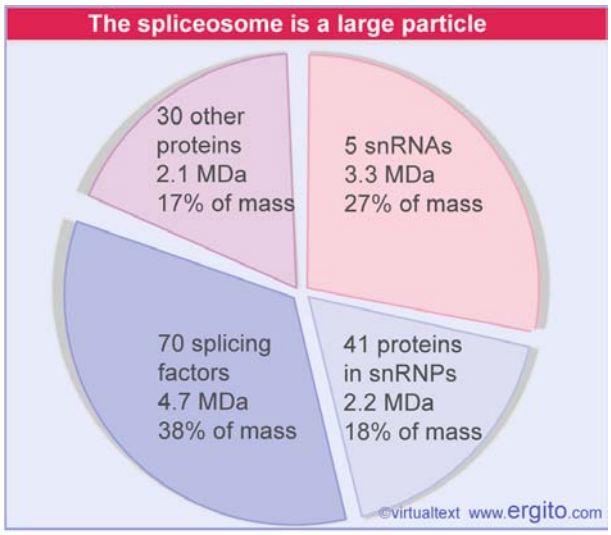

Figure 24.8 The spliceosome is ~12 MDa. 5 snRNAPs account for almost half of the mass. The remaining proteins include known splicing factors and also proteins that are involved in other stages of gene expression.

The spliceosome forms on the intact precursor RNA and passes through an intermediate state in which it contains the individual 5′ exon linear molecule and the right lariat-intron-exon. Little spliced product is found in the complex, which suggests that it is usually released immediately following the cleavage of the 3′ site and ligation of the exons.

We may think of the snRNP particles as being involved in building the structure of the spliceosome. Like the ribosome, the spliceosome depends on RNA-RNA interactions as well as protein-RNA and protein-protein interactions. Some of the reactions involving the snRNPs require their RNAs to base pair directly with sequences in the RNA being spliced; other reactions require recognition between snRNPs or between their proteins and other components of the spliceosome.

The importance of snRNA molecules can be tested directly in yeast by making mutations in their genes. Mutations in 5 snRNA genes are lethal and prevent splicing. All of the snRNAs involved in splicing can be recognized in conserved forms in animal, bird, and insect cells. The corresponding RNAs in yeast are often rather larger, but conserved regions include features that are similar to the snRNAs of higher eukaryotes.

The snRNPs involved in splicing are U1, U2, U5, U4, and U6. They are named according to the snRNAs that are present. Each snRNP contains a single snRNA and several (<20) proteins. The U4 and U6 snRNPs are usually found as a single (U4/U6) particle. A common structural core for each snRNP consists of a group of 8 proteins, all of which are recognized by an autoimmune antiserum called **anti-Sm**; conserved sequences in the proteins form the target for the antibodies. The other proteins in each snRNP are unique to it. The Sm proteins bind to the conserved sequence $PuAU_{3-6}Gpu$, which is present in all snRNAs except U6. The U6 snRNP contains instead a set of Sm-like (Lsm) proteins. The Sm proteins must be involved in the autoimmune reaction, although their relationship to the phenotype of the autoimmune disease is not clear (for review see 244; 245; 247).

Some of the proteins in the snRNPs may be involved directly in splicing; others may be required in structural roles or just for assembly or interactions between the snRNP particles. About one third of the proteins involved in splicing are components of the snRNPs. Increasing evidence for a direct role of RNA in the splicing reaction suggests that relatively few of the splicing factors play a direct role in catalysis; most are involved in structural or assembly roles.

Last updated on 1-3-2003

Reviews

244. Maniatis, T. and Reed, R. (1987). *The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing*. Nature 325, 673-678.

245. Guthrie, C. and Patterson, B. (1988). *Spliceosomal snRNAs*. Annu. Rev. Genet. 22, 387-419.

247. Guthrie, C. (1991). *Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein*. Science 253, 157-163.

References

719. Grabowski, P. J., Seiler, S. R., and Sharp, P. A. (1985). *A multicomponent complex is involved in the splicing of messenger RNA precursors*. Cell 42, 345-353.

3210. Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). *Comprehensive proteomic analysis of the human spliceosome.* Nature 419, 182-185.

RNA SPLICING AND PROCESSING

5.24.6 U1 snRNP initiates splicing

Key Terms

An **SR protein** has a variable length of an Arg-Ser-rich region and is involved in splicing.

Key Concepts

- U1 snRNP initiates splicing by binding to the 5′ splice site by means of an RNA-RNA pairing reaction.
- The E complex contains U1 snRNP bound at the 5′ splice site, the protein U2AF bound to a pyrimidine tract between the branch site and the 3′ splice site, and SR proteins connecting U1 snRNP to U2AF.

Splicing can be broadly divided into two stages:

- First the consensus sequences at the 5′ splice site, branch sequence, and adjacent pyrimidine tract are recognized. A complex assembles that contains all of the splicing components.
- Then the cleavage and ligation reactions change the structure of the substrate RNA. Components of the complex are released or reorganized as it proceeds through the splicing reactions.

The important point is that all of the splicing components are assembled and have assured that the splice sites are available before any irreversible change is made to the RNA (for review see 3239).

Recognition of the consensus sequences involves both RNAs and proteins. Certain snRNAs have sequences that are complementary to the consensus sequences or to one another, and base pairing between snRNA and pre-mRNA, or between snRNAs, plays an important role in splicing.

The human U1 snRNP contains 8 proteins as well as the RNA. The secondary structure of the U1 snRNA is drawn in **Figure 24.9**. It contains several domains. The Sm-binding site is required for interaction with the common snRNP proteins. Domains identified by the individual stem-loop structures provide binding sites for proteins that are unique to U1 snRNP.

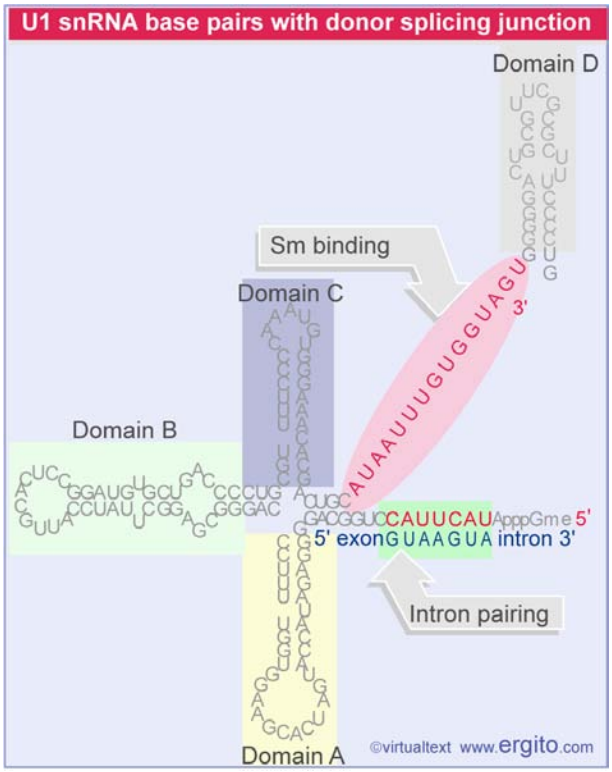

Figure 24.9 U1 snRNA has a base paired structure that creates several domains. The 5′ end remains single stranded and can base pair with the 5′ splicing site.

Binding of U1 snRNP to the 5′ splice site is the first step in splicing. The recruitment of U1 snRNP involves an interaction between one of its proteins (U1-70k) and the protein ASF/SF2 (a general splicing factor in the SR class: see below). U1 snRNA base pairs with the 5′ site by means of a single-stranded region at its 5′ –terminus which usually includes a stretch of 4-6 bases that is complementary with the splice site.

Mutations in the 5′ splice site and U1 snRNA can be used to test directly whether pairing between them is necessary. The results of such an experiment are illustrated in **Figure 24.10**. The wild-type sequence of the splice site of the 12S adenovirus pre-mRNA pairs at 5 out of 6 positions with U1 snRNA. A mutant in the 12S RNA that cannot be spliced has two sequence changes; the GG residues at positions 5-6 in the intron are changed to AU. The mutation changes the pattern of base pairing between U1 snRNA and the 5′ splice site, although it does not alter the *overall* extent of pairing (because complementarity is lost at one position and gained at the other). The effect on splicing suggests that the base-pairing interaction is important.

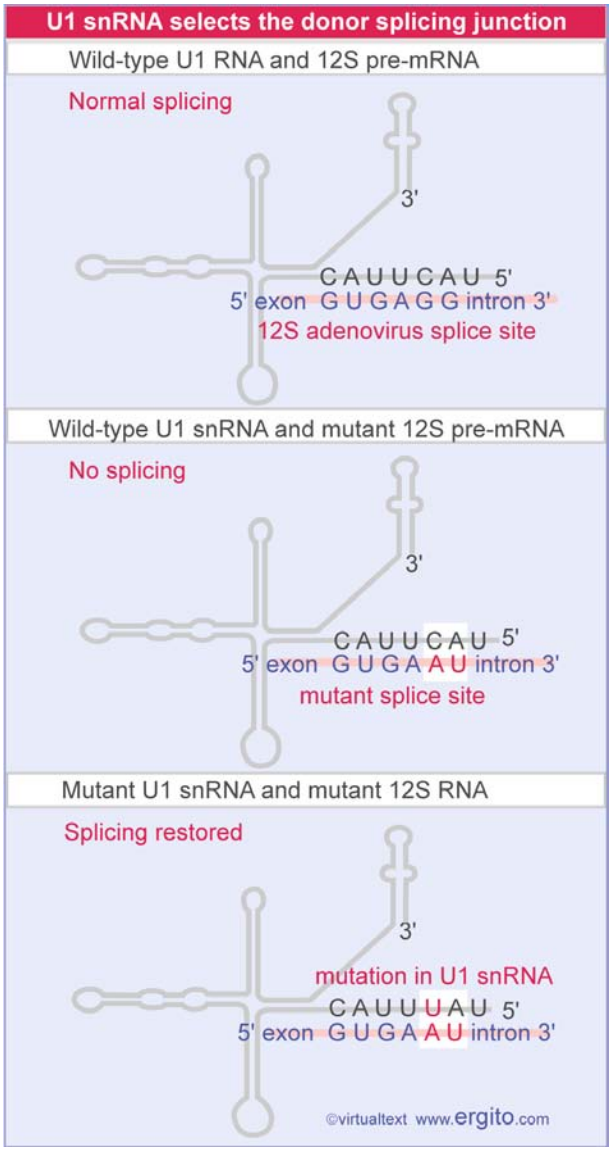

Figure 24.10 Mutations that abolish function of the 5′ splicing site can be suppressed by compensating mutations in U1 snRNA that restore base pairing.

When a mutation is introduced into U1 snRNA that restores pairing at position 5, normal splicing is regained. Other cases in which corresponding mutations are made in U1 snRNA to see whether they can suppress the mutation in the splice site suggests the general rule: complementarity between U1 snRNA and the 5′ splice site is necessary for splicing, but the efficiency of splicing is not determined solely by the number of base pairs that can form (714). The pairing reaction is stabilized by the proteins of the U1 snRNP (3246).

Figure 24.11 shows the early stages of splicing. The first complex formed during splicing is the E (early presplicing) complex, which contains U1 snRNP, the splicing factor U2AF, and members of a family called **SR proteins**, which comprise an important group of splicing factors and regulators. They take their name from the

presence of an Arg-Ser-rich region that is variable in length. SR proteins interact with one another via their Arg-Ser-rich regions. They also bind to RNA. They are an essential component of the spliceosome, forming a framework on the RNA substrate. They connect U2AF to U1 (see **Figure 24.12**). The E complex is sometimes called the commitment complex, because its formation identifies a pre-mRNA as a substrate for formation of the splicing complex.

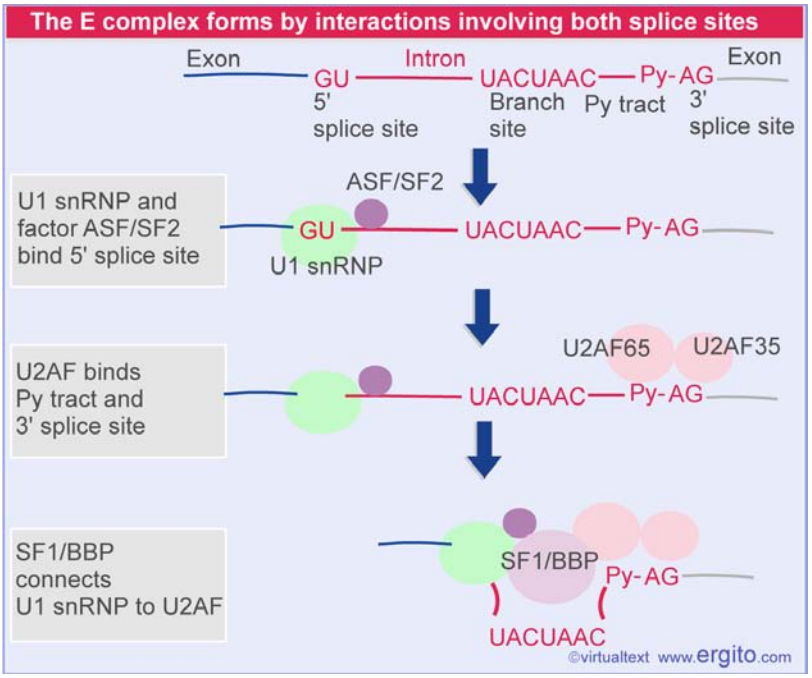

Figure 24.11 The commitment (E) complex forms by the successive addition of U1 snRNP to the 5′ splice site, U2AF to the pyrimidine tract/3′ splice site, and the bridging protein SF1/BBP.

In the E complex, U2AF is bound to the region between the branch site and the 3′ splice site. The name of U2AF reflects its original isolation as the U2 auxiliary factor. In most organisms, it has a large subunit (U2AF65) that contacts a pyrimidine tract downstream of the branch site, while a small subunit (U2AF35) directly contacts the dinucleotide AG at the 3′ splice site (3247; 3248; 3249). In *S. cerevisiae*, this function is filled by the protein Mud2, which is a counterpart of U2AF65, and binds only to the pyrimidine tract. This marks a difference in the mechanism of splicing between *S. cerevisiae* and other organisms. In the yeast, the 3′ splice site is not involved in the early stages of forming the splicing complex, but in all other known cases, it is required.

Another splicing factor, called SF1 in mammals and BBP in yeast, connects U2AF/Mud2 to the U1 snRNP bound at the 5′ splice site (3250; 3251). Complex formation is enhanced by the cooperative reactions of the two proteins; SF1 and U2AF (or BBP and Mud2) bind together to the RNA substrate ~10× more effectively than either alone. This interaction is probably responsible for making the first connection between the two splice sites across the intron.

The E complex is converted to the A complex when U2 snRNP binds to the branch site. Both U1 snRNP and U2AF/Mud2 are needed for U2 binding. The U2 snRNA

includes sequences complementary to the branch site. A sequence near the 5′ end of the snRNA base pairs with the branch sequence in the intron. In yeast this typically involves formation of a duplex with the UACUAAC box (see **Figure 24.14**). Several proteins of the U2 snRNP are bound to the substrate RNA just upstream of the branch site. The addition of U2 snRNP to the E complex generates the A presplicing complex. The binding of U2 snRNP requires ATP hydrolysis, and commits a pre-mRNA to the splicing pathway (1560; 718).

Last updated on 1-8-2003

Reviews

3239. Brow, D. A. (2002). *Allosteric cascade of spliceosome activation.* Annu. Rev. Genet. 36, 333-360.

References

714. Zhuang, Y. and Weiner, A. M. (1986). *A compensatory base change in U1 snRNA suppresses a 5 splice site mutation.* Cell 46, 827-835.

718. Burgess, S., Couto, J. R., and Guthrie, C. (1990). *A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing.* Cell 60, 705-717.

1560. Parker, R., Siliciano, P. G., and Guthrie, C. (1987). *Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA.* Cell 49, 229-239.

3246. Zhang, D. and Rosbash, M. (1999). *Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex.* Genes Dev. 13, 581-592.

3247. Zamore, P. D. and Green, M. R. (1989). *Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor.* Proc. Natl. Acad. Sci. USA 86, 9243-9247.

3248. Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999). *Functional recognition of the 3′ splice site AG by the splicing factor U2AF35.* Nature 402, 832-835.

3249. Singh, R., Valca¡rcel, J., and Green, M. R. (1995). *Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins.* Science 268, 1173-1176.

3250. Abovich, N. and Rosbash, M. (1997). *Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals.* Cell 89, 403-412.

3251. Berglund, J. A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). *The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC.* Cell 89, 781-787.

RNA SPLICING AND PROCESSING

5.24.7 The E complex can be formed by intron definition or exon definition

Key Terms

Intron definition describes the process when a pair of splicing sites are recognized by interactions involving only the 5′ site and the branchpoint/3′ site.

Exon definition describes the process when a pair of splicing sites are recognized by interactions involving the 5′ site of the intron and also the 5′ of the next intron downstream.

Key Concepts

- The direct way of forming an E complex is for U1 snRNP to bind at the 5′ splice site and U2AF to bind at a pyrimidine tract between the branch site and the 3′ splice site.
- Another possibility is for the complex to form between U2AF at the pyrimidine tract and U1 snRNP at a downstream 5′ splice site.
- The E complex is converted to the A complex when U2 snRNP binds at the branch site.
- If an E complex forms using a downstream 5′ splice site, this splice site is replaced by the appropriate upstream 5′ splice site when the E complex is converted to the A complex.
- Weak 3′ splice sites may require a splicing enhancer located in the exon downstream to bind SR proteins directly.

There is more than one way to form the E complex. **Figure 24.12** illustrates some possibilities. The most direct reaction is for both splice sites to be recognized across the intron. The presence of U1 snRNP at the 5′ splice site is necessary for U2AF to bind at the pyrimidine tract downstream of the branch site, making it possible that the 5′ and 3′ ends of the intron are brought together in this complex. The E complex is converted to the A complex when U2 snRNP binds at the branch site. *The basic feature of this route for splicing is that the two splice sites are recognized without requiring any sequences outside of the intron.* This process is called **intron definition**.

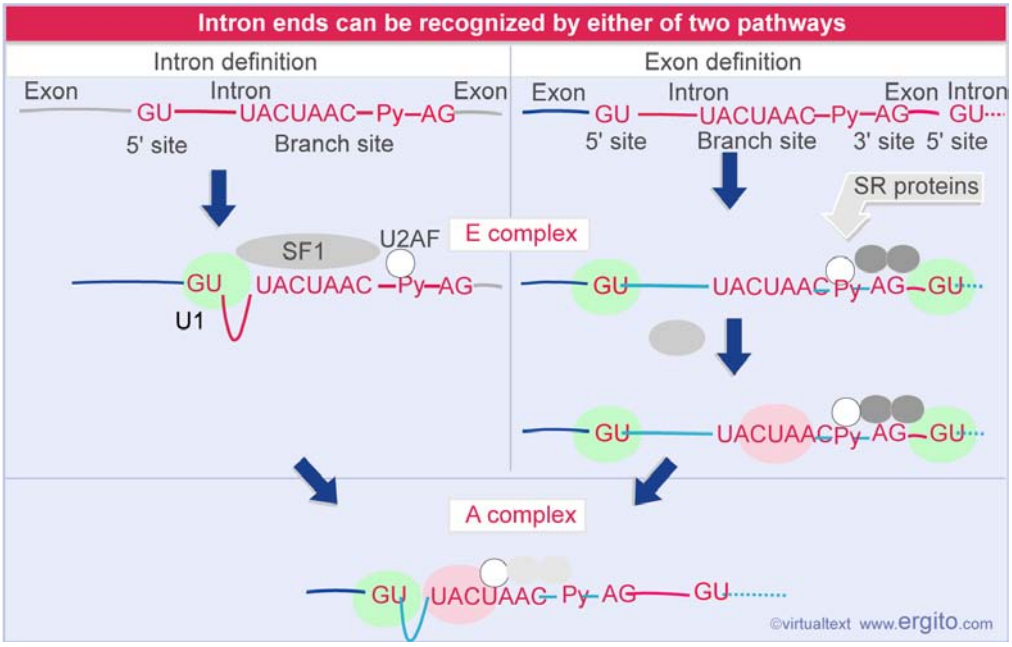

Figure 24.12 There may be multiple routes for initial recognition of 5′ and 3′ splice sites.

In an extreme case, the SR proteins may enable U2AF/U2 snRNP to bind *in vitro* in the absence of U1, raising the possibility that there could be a U1-independent pathway for splicing.

An alternative route to form the spliceosome may be followed when the introns are long and the splice sites are weak. As shown on the right of the figure, the 5′ splice site is recognized by U1 snRNA in the usual way. However, the 3′ splice site is recognized as part of a complex that forms across the *next exon*, in which the next 5′ splice site is also bound by U1 snRNA. This U1 snRNA is connected by SR proteins to the U2AF at the pyrimidine tract. When U2 snRNP joins to generate the A complex, there is a rearrangement, in which the correct (leftmost) 5′ splice site displaces the downstream 5′ splice site in the complex. The important feature of this route for splicing is that sequences downstream of the intron itself are required. Usually these sequences include the next 5′ splice site. This process is called **exon definition**. This mechanism is not universal: neither SR proteins nor exon definition are found in *S. cerevisiae*.

"Weak" 3′ splice sites do not bind U2AF and U2 snRNP effectively. Additional sequences are needed to bind the SR proteins, which assist U2AF in binding to the pyrimidine tract. Such sequences are called "splicing enhancers," and they are most commonly found in the exon downstream of the 3′ splice site (734).

References

734. Bruzik, J. P. and Maniatis, T. (1995). *Enhancer-dependent interaction between 5′ and 3′ splice sites in trans*. Proc. Natl. Acad. Sci. USA 92, 7056-7059.

RNA SPLICING AND PROCESSING

5.24.8 5 snRNPs form the spliceosome

Key Concepts

- Binding of U5 and U4/U6 snRNPs converts the A complex to the B1 spliceosome, which contains all the components necessary for splicing.
- The spliceosome passes through a series of further complexes as splicing proceeds.
- Release of U1 snRNP allows U6 snRNA to interact with the 5′ splice site, and converts the B1 spliceosome to the B2 spliceosome.
- When U4 dissociates from U6 snRNP, U6 snRNA can pair with U2 snRNA to form the catalytic active site.

Following formation of the E complex, the other snRNPs and factors involved in splicing associate with the complex in a defined order. **Figure 24.13** shows the components of the complexes that can be identified as the reaction proceeds.

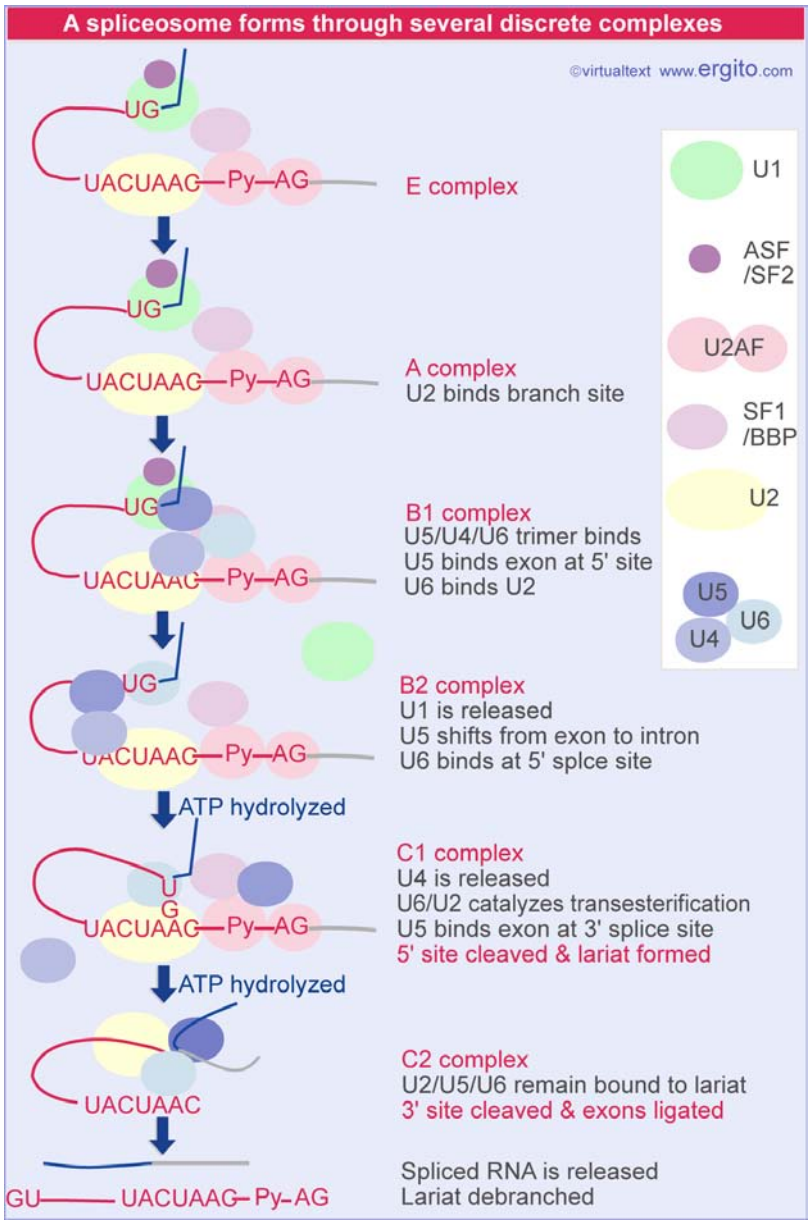

Figure 24.13 The splicing reaction proceeds through discrete stages in which spliceosome formation involves the interaction of components that recognize the consensus sequences.

The B1 complex is formed when a trimer containing the U5 and U4/U6 snRNPs binds to the A complex containing U1 and U2 snRNPs. This complex is regarded as a spliceosome, since it contains the components needed for the splicing reaction. It is converted to the B2 complex after U1 is released. The dissociation of U1 is necessary to allow other components to come into juxtaposition with the 5′ splice site, most notably U6 snRNA. At this point U5 snRNA changes its position; initially it is close to exon sequences at the 5′ splice site, but it shifts to the vicinity of the intron sequences (722).

The catalytic reaction is triggered by the release of U4; this requires hydrolysis of

ATP. The role of U4 snRNA may be to sequester U6 snRNA until it is needed. **Figure 24.14** shows the changes that occur in the base pairing interactions between snRNAs during splicing. In the U6/U4 snRNP, a continuous length of 26 bases of U6 is paired with two separated regions of U4. When U4 dissociates, the region in U6 that is released becomes free to take up another structure. The first part of it pairs with U2; the second part forms an intramolecular hairpin. The interaction between U4 and U6 is mutually incompatible with the interaction between U2 and U6, so the release of U4 controls the ability of the spliceosome to proceed (721; for review see 252).

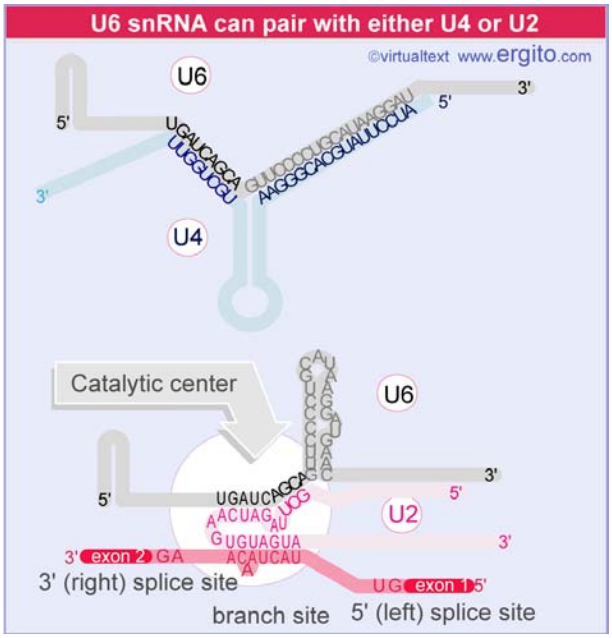

Figure 24.14 U6-U4 pairing is incompatible with U6-U2 pairing. When U6 joins the spliceosome it is paired with U4. Release of U4 allows a conformational change in U6; one part of the released sequence forms a hairpin (dark grey), and the other part (black) pairs with U2. Because an adjacent region of U2 is already paired with the branch site, this brings U6 into juxtaposition with the branch. Note that the substrate RNA is reversed from the usual orientation and is shown 3′-5′.

Although for clarity the figure shows the RNA substrate in extended form, the 5′ splice site is actually close to the U6 sequence immediately on the 5′ side of the stretch bound to U2. This sequence in U6 snRNA pairs with sequences in the intron just downstream of the conserved GU at the 5′ splice site (mutations that enhance such pairing improve the efficiency of splicing).

So several pairing reactions between snRNAs and the substrate RNA occur in the course of splicing. They are summarized in **Figure 24.15**. The snRNPs have sequences that pair with the substrate and with one another. They also have single-stranded regions in loops that are in close proximity to sequences in the substrate, and which play an important role, as judged by the ability of mutations in the loops to block splicing.

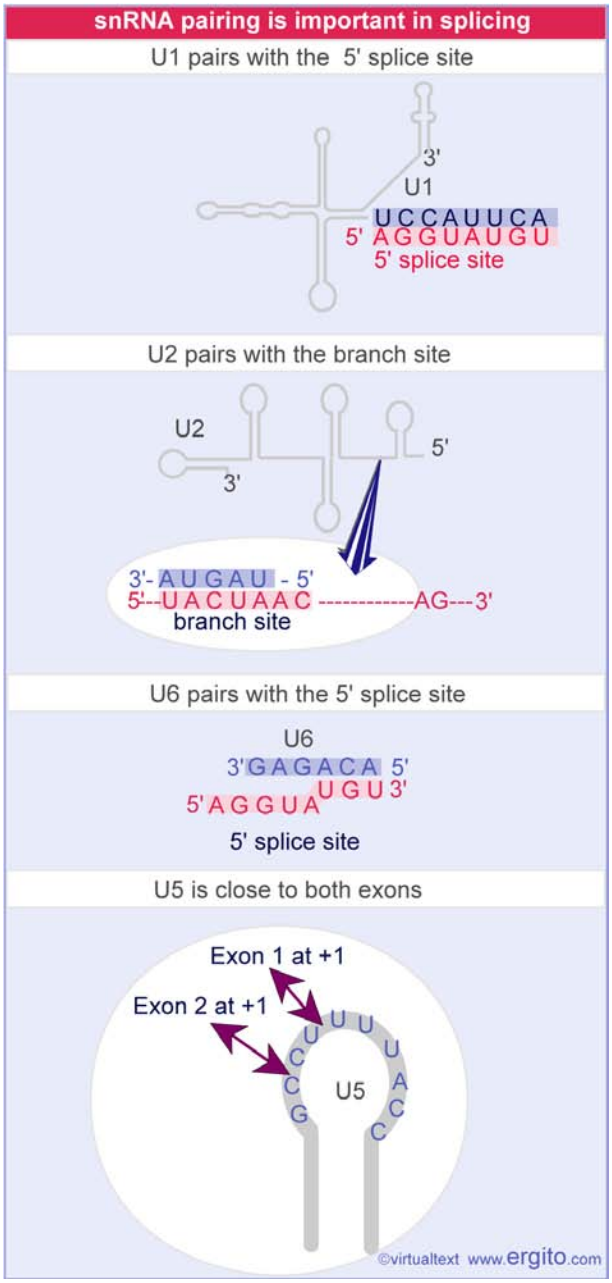

Figure 24.15 Splicing utilizes a series of base pairing reactions between snRNAs and splice sites.

The base pairing between U2 and the branch point, and between U2 and U6, creates a structure that resembles the active center of group II self-splicing introns (see **Figure 24.20**). This suggests the possibility that the catalytic component could comprise an RNA structure generated by the U2-U6 interaction. U6 is paired with the 5′ splice site, and crosslinking experiments show that a loop in U5 snRNA is immediately adjacent to the first base positions in both exons. But although we can define the proximities of the substrate (5′ splice site and branch site) and snurps (U2 and U6) at the catalytic center (as shown in **Figure 24.14**), the components that undertake the transesterifications have not been directly identified (723; 724; 725).

The formation of the lariat at the branch site is responsible for determining the use of the 3′ splice site, since the 3′ consensus sequence nearest to the 3′ side of the branch becomes the target for the second transesterification. The second splicing reaction follows rapidly. Binding of U5 snRNP to the 3′ splice site is needed for this reaction, but there is no evidence for a base pairing reaction.

The important conclusion suggested by these results is that *the snRNA components of the splicing apparatus interact both among themselves and with the substrate RNA by means of base pairing interactions, and these interactions allow for changes in structure that may bring reacting groups into apposition and may even create catalytic centers.* Furthermore, the conformational changes in the snRNAs are reversible; for example, U6 snRNA is not used up in a splicing reaction, and at completion must be released from U2, so that it can reform the duplex structure with U4 to undertake another cycle of splicing.

We have described individual reactions in which each snRNP participates, but as might be expected from a complex series of reactions, any particular snRNP may play more than one role in splicing. So the ability of U1 snRNP to promote binding of U2 snRNP to the branch site is independent of its ability to bind to the 5′ splice site. Similarly, different regions of U2 snRNA can be defined that are needed to bind to the branch site and to interact with other splicing components.

An extensive mutational analysis has been undertaken in yeast to identify both the RNA and protein components of the spliceosome. Mutations in genes needed for splicing are identified by the accumulation of unspliced precursors. A series of loci that identify genes potentially coding for proteins involved in splicing were originally called *RNA*, but are now known as *PRP* mutants (for pre-RNA processing). Several of the products of these genes have motifs that identify them as RNA-binding proteins, and some appear to be related to a family of ATP-dependent RNA helicases. We suppose that, in addition to RNA-RNA interactions, protein-RNA interactions are important in creating or releasing structures in the pre-mRNA or snRNA components of the spliceosomes.

Some of the PRP proteins are components of snRNP particles, but others function as independent factors. One interesting example is PRP16, a helicase that hydrolyzes ATP, and associates transiently with the spliceosome to participate in the second catalytic step. Another example is PRP22, another ATP-dependent helicase, which is required to release the mature mRNA from the spliceosome. The conservation of bonds during the splicing reaction means that input of energy is not required to drive bond formation *per se*, which implies that the ATP hydrolysis is required for other purposes. The use of ATP by PRP16 and PRP22 may be examples of a more general phenomenon: the use of ATP hydrolysis to drive conformational changes that are needed to proceed through splicing (for review see 254).

Reviews

252. Madhani, H. D. and Guthrie, C. (1994). *Dynamic RNA-RNA interactions in the spliceosome.* Annu. Rev. Genet. 28, 1-26.

254. Krämer, A. (1996). *The structure and function of proteins involved in mammalian pre-mRNA splicing*. Annu. Rev. Biochem. 65, 367-409.

References

721. Lamond, A. I. (1988). *Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein*. Proc. Natl. Acad. Sci. USA 85, 411-415.

722. Newman, A. and Norman, C. (1991). *Mutations in yeast U5 snRNA alter the specificity of 5′ splice site cleavage*. Cell 65, 115-123.

723. Madhani, H. D. and Guthrie, C. (1992). *A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome*. Cell 71, 803-817.

724. Lesser, C. F. and Guthrie, C. (1993). *Mutations in U6 snRNA that alter splice site specificity: implications for the active site*. Science 262, 1982-1988.

725. Sontheimer, E. J. and Steitz, J. A. (1993). *The U5 and U6 small nuclear RNAs as active site components of the spliceosome*. Science 262, 1989-1996.

RNA SPLICING AND PROCESSING

5.24.9 An alternative splicing apparatus uses different snRNPs

Key Concepts

- An alternative splicing pathway uses another set of snRNPs that comprise the U12 spliceosome.
- The target introns are defined by longer consensus sequences at the splice junctions, but usually include the same GU-AG junctions.
- Some introns have the splice junctions AU-AC, including some that are U1-dependent and some that are U12-dependent.

GU-AG introns comprise the vast majority (>98% of splicing junctions in the human genome). <1% use the related junctions GC-AG. And then there is a minor class of introns marked by the ends AU-AC (comprising 0.1% of introns). The first of these introns to be discovered required an alternative splicing apparatus, called the U12 spliceosome, consisting of U11 and U12 (related to U1 and U2, respectively), a U5 variant, and the $U4_{atac}$ and $U6_{atac}$ snRNAs. The splicing reaction is essentially similar to that at GU-AG introns, and the snRNAs play analogous roles. Whether there are differences in the protein components of this apparatus is not known.

It now turns out that the dependence on the type of spliceosome is also influenced by sequences in the intron, so that there are some AU-AC introns spliced by U2-type spliceosomes, and some GU-AG introns spliced by U12-type spliceosomes. A strong consensus sequence at the left end defines the U12-dependent type of intron: 5′ $^{G}_{A}$UAUCCUUU# ...PyA^{G}_{C} 3′. In fact, most U12-dependent introns have the GU.. ...AG termini. In addition, they have a highly conserved branch point, UCCUUPuAPy, which pairs with U12. For this reason, the term U12-dependent intron is used rather than AU-AC intron.

The two types of introns coexist in a variety of genomes, and in some cases are found in the same gene. U12-dependent introns tend to be flanked by U2-dependent introns. What is known about the phylogeny of these introns suggests that AU-AC U12-dependent introns may once have been more common, but tend to be converted to GU-AG termini, and to U2-dependence, in the course of evolution (726; 727; 728). The common evolution of the systems is emphasized by the fact that they use analogous sets of base pairing between the snRNAs and with the substrate pre-mRNA.

The involvement of snRNPs in splicing is only one example of their involvement in RNA processing reactions. snRNPs are required for several reactions in the processing of nuclear RNA to mature rRNAs. Especially in view of the demonstration that group I introns are self-splicing, and that the RNA of ribonuclease P has catalytic activity (as discussed in *Molecular Biology 5.26 Catalytic RNA*), it is plausible to think that RNA-RNA reactions are important in many RNA processing

events.

Last updated on 2-16-2001

References

726. Tarn, W.-Y. and Steitz, J. (1996). *A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class AT-AC intron in vitro.* Cell 84, 801-811.

727. Dietrich, R. C., Incorvaia, R., and Padgett, R. A. (1997). *Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns Molecular*. Mol. Cell 1, 151-160.

728. Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998). *Evolutionary fates and origins of U12-type introns*. Mol. Cell 2, 773-785.

RNA SPLICING AND PROCESSING

5.24.10 Splicing is connected to export of mRNA

Key Concepts

- The REF proteins bind to splicing junctions by associating with the spliceosome.
- After splicing, they remain attached to the RNA at the exon-exon junction.
- They interact with the transport protein TAP/Mex that exports the RNA through the nuclear pore.

After it has been synthesized and processed, mRNA is exported from the nucleus to the cytoplasm in the form of a ribonucleoprotein complex. The proteins that are responsible for transport "shuttle" between the nucleus and cytoplasm, remaining in the compartment only briefly (see *Molecular Biology 2.8.28 Transport receptors carry cargo proteins through the pore*). One important question is how these proteins recognize their RNA substrates, and what ensures that only fully processed mRNAs are exported. Part of the answer may lie in the relative timing of events: spliceosomes may form to remove introns before transcription has been completed. However, there may also be a direct connection between splicing and export.

Introns may prevent export of mRNA because they are associated with the splicing apparatus. The spliceosome also may provide the initial point of contact for the export apparatus. **Figure 24.16** shows a model in which a protein complex binds to the RNA via the splicing apparatus. The complex consists of >9 proteins and is called the EJC (exon junction complex).

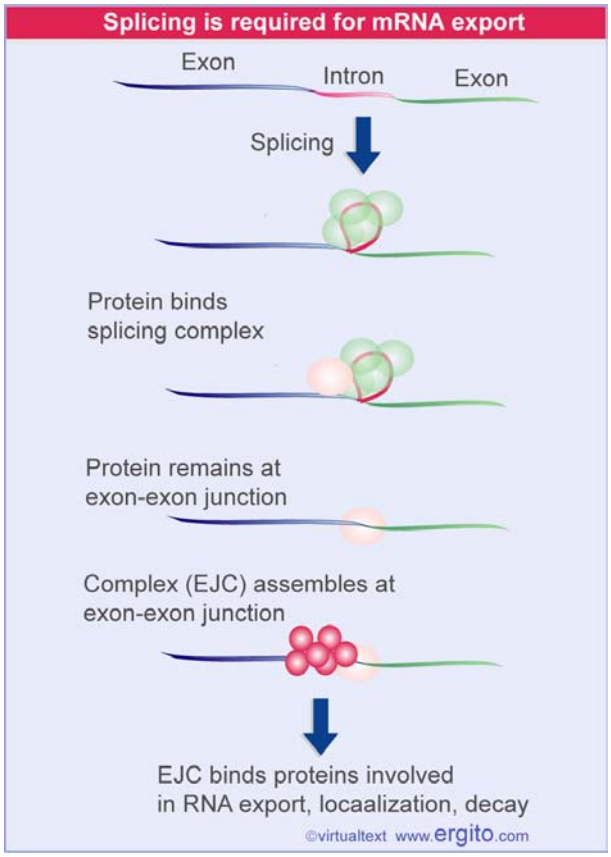

Figure 24.16 The EJC (exon junction complex) binds to RNA by recognizing the splicing complex.

The EJC is involved in several functions of spliced mRNAs (for review see 3428). Some of the proteins of the EJC are directly involved in these functions, and others recruit additional proteins for particular functions. The first contact in assembling the EJC is made with one of the splicing factors (2072; 2073; 2074). Then after splicing, the EJC remains attached to the mRNA just upstream of the exon-exon junction (2071; 2075; 3232; 3233). The EJC is not associated with RNAs transcribed from genes that lack introns, so its involvement in the process is unique for spliced products.

If introns are deleted from a gene, its RNA product is exported much more slowly to the cytoplasm (2069). This suggests that the intron may provide a signal for attachment of the export apparatus. We can now account for this phenomenon in terms of a series of protein interactions, as shown in **Figure 24.17**. The EJC includes a group of proteins called the REF family (the best characterized member is called Aly) (2070). The REF proteins in turn interact with a transport protein (variously called TAP and Mex) which has direct responsibility for interaction with the nuclear pore (for review see 2422).

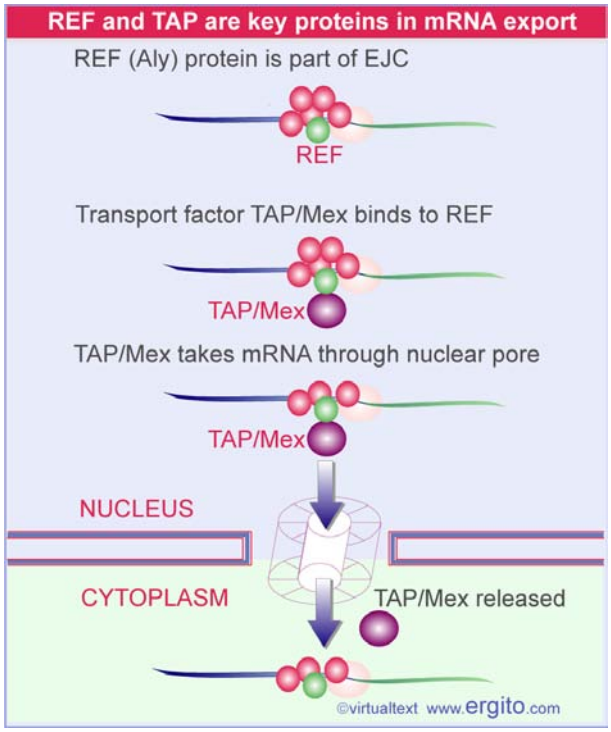

Figure 24.17 A REF protein binds to a splicing factor and remains with the spliced RNA product. REF binds to an export factor that binds to the nuclear pore.

A similar system may be used to identify a spliced RNA so that nonsense mutations prior to the last exon trigger its degradation in the cytoplasm (see *Molecular Biology 2.5.14 Nonsense mutations trigger a surveillance system*).

Last updated on 1-6-2003

Reviews

2422. Reed, R. and Hurt, E. (2002). *A conserved mRNA export machinery coupled to pre-mRNA splicing.* Cell 108, 523-531.

3428. Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). *Messenger-RNA-binding proteins and the messages they carry.* Nat. Rev. Mol. Cell Biol. 3, 195-205.

References

2069. Luo, M. J. and Reed, R. (1999). *Splicing is required for rapid and efficient mRNA export in metazoans.* Proc. Natl. Acad. Sci. USA 96, 14937-14942.

2070. Rodrigues, J. P., Rode, M., Gatfield, D., Blencowe, B., Blencowe, M., and Izaurralde, E. (2001). *REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus.* Proc. Natl. Acad. Sci. USA 98, 1030-1035.

2071. Le Hir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000). *The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions.* EMBO J. 19, 6860-6869.

2072. Kataoka, N., Yong, J., Kim, V. N., Velazquez, F., Perkinson, R. A., Wang, F., and Dreyfuss, G. (2000). *Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm.* Mol. Cell 6, 673-682.

2073. Luo, M. L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. (2001). *Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly.* Nature 413, 644-647.

2074. Strasser, K. and Hurt, E. (2001). *Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p.* Nature 413, 648-652.

2075. Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). *The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans.* Nature 407, 401-405.

3232. Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M. J. (2001). *The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay.* EMBO J. 20, 4987-4997.

3233. Reichert, V. L., Le Hir, H., Jurica, M. S., and Moore, M. J. (2002). *5′ exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly.* Genes Dev. 16, 2778-2791.

RNA SPLICING AND PROCESSING

5.24.11 Group II introns autosplice via lariat formation

Key Terms

Autosplicing (Self-splicing) describes the ability of an intron to excise itself from an RNA by a catalytic action that depends only on the sequence of RNA in the intron.

Key Concepts

- Group II introns excise themselves from RNA by an autocatalytic splicing event.
- The splice junctions and mechanism of splicing of group II introns are similar to splicing of nuclear introns.
- A group II intron folds into a secondary structure that generates a catalytic site resembling the structure of U6-U2-nuclear intron.

Introns in protein-coding genes (in fact, in all genes except nuclear tRNA-coding genes) can be divided into three general classes. Nuclear pre-mRNA introns are identified only by the possession of the GU...AG dinucleotides at the 5′ and 3′ ends and the branch site/pyrimidine tract near the 3′ end. They do not show any common features of secondary structure. Group I and group II introns are found in organelles and in bacteria. (Group I introns are found also in the nucleus in lower eukaryotes.) Group I and group II introns are classified according to their internal organization. Each can be folded into a typical type of secondary structure.

The group I and group II introns have the remarkable ability to excise themselves from an RNA. This is called **autosplicing**. Group I introns are more common than group II introns. There is little relationship between the two classes, but in each case the RNA can perform the splicing reaction *in vitro* by itself, without requiring enzymatic activities provided by proteins; however, proteins are almost certainly required *in vivo* to assist with folding (see *Molecular Biology 5.26 Catalytic RNA*).

Figure 24.18 shows that three classes of introns are excised by two successive transesterifications (shown previously for nuclear introns in **Figure 24.6**). In the first reaction, the 5′ exon-intron junction is attacked by a free hydroxyl group (provided by an internal 2′–OH position in nuclear and group II introns, and by a free guanine nucleotide in group I introns). In the second reaction, the free 3′–OH at the end of the released exon in turn attacks the 3′ intron-exon junction.

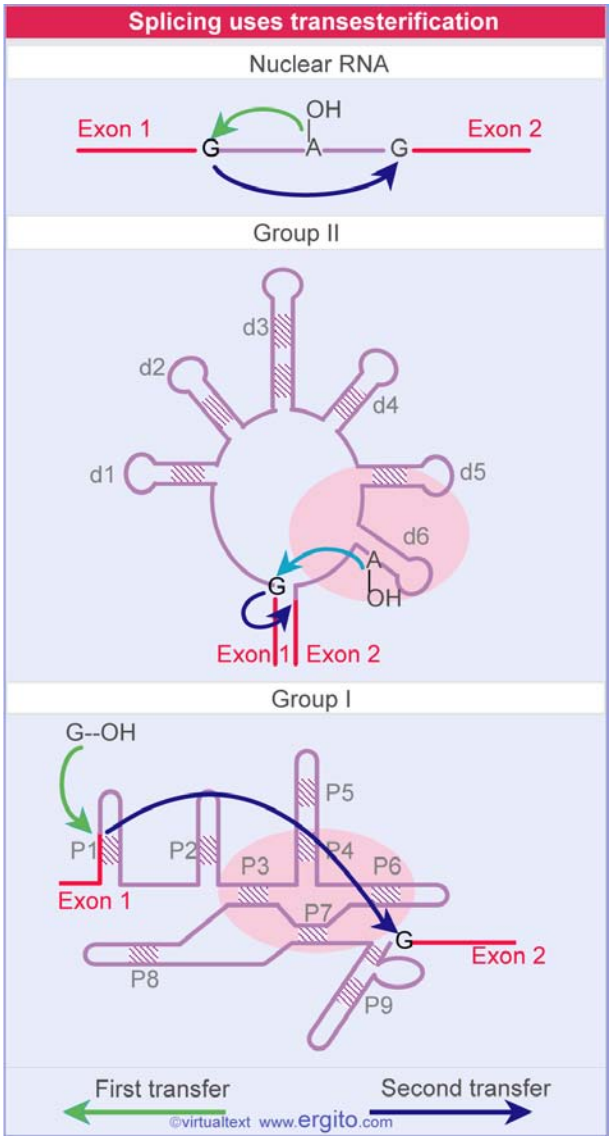

Figure 24.18 Three classes of splicing reactions proceed by two transesterifications. First, a free OH group attacks the exon 1–intron junction. Second, the OH created at the end of exon 1 attacks the intron–exon 2 junction.

There are parallels between group II introns and pre-mRNA splicing. Group II mitochondrial introns are excised by the same mechanism as nuclear pre-mRNAs, via a lariat that is held together by a 5′–2′ bond. An example of a lariat produced by splicing a group II intron is shown in **Figure 24.19**. When an isolated group II RNA is incubated *in vitro* in the absence of additional components, it is able to perform the splicing reaction. This means that the two transesterification reactions shown in **Figure 24.18** can be performed by the group II intron RNA sequence itself. Because the number of phosphodiester bonds is conserved in the reaction, an external supply of energy is not required; this could have been an important feature in the evolution of splicing (for review see 260).

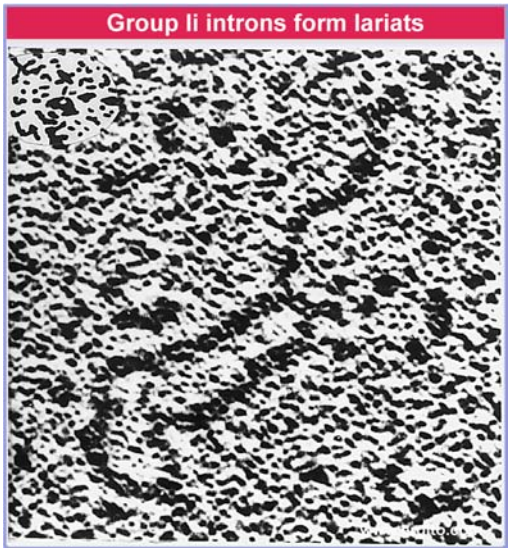

Figure 24.19 Splicing releases a mitochondrial group II intron in the form of a stable lariat. Photograph kindly provided by Leslie Grivell and Annika Arnberg.

A group II intron forms into a secondary structure that contains several domains formed by base-paired stems and single-stranded loops. Domain 5 is separated by 2 bases from domain 6, which contains an A residue that donates the 2′–OH group for the first transesterification. This constitutes a catalytic domain in the RNA. **Figure 24.20** compares this secondary structure with the structure formed by the combination of U6 with U2 and of U2 with the branch site. The similarity suggests that U6 may have a catalytic role.

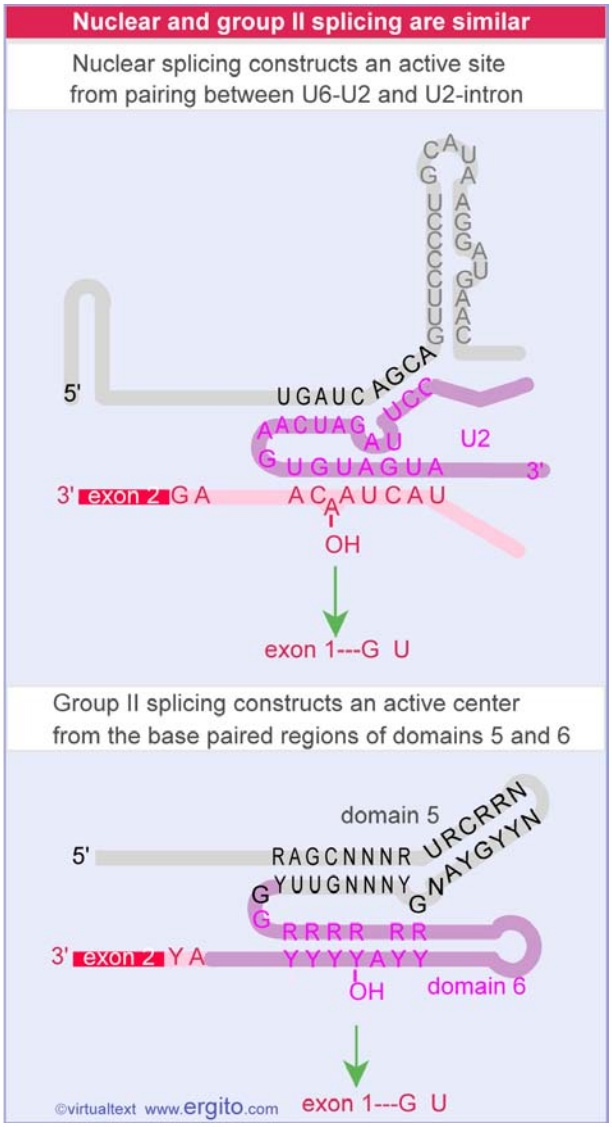

Figure 24.20 Nuclear splicing and group II splicing involve the formation of similar secondary structures. The sequences are more specific in nuclear splicing; group II splicing uses positions that may be occupied by either purine (R) or either pyrimidine (Y).

The features of group II splicing suggest that splicing evolved from an autocatalytic reaction undertaken by an individual RNA molecule, in which it accomplished a controlled deletion of an internal sequence. Probably such a reaction requires the RNA to fold into a specific conformation, or series of conformations, and would occur exclusively in *cis* conformation.

The ability of group II introns to remove themselves by an autocatalytic splicing event stands in great contrast to the requirement of nuclear introns for a complex splicing apparatus. We may regard the snRNAs of the spliceosome as compensating for the lack of sequence information in the intron, and providing the information required to form particular structures in RNA. The functions of the snRNAs may

have evolved from the original autocatalytic system. These snRNAs act in *trans* upon the substrate pre-mRNA; we might imagine that the ability of U1 to pair with the 5′ splice site, or of U2 to pair with the branch sequence, replaced a similar reaction that required the relevant sequence to be carried by the intron. So the snRNAs may undergo reactions with the pre-mRNA substrate and with one another that have substituted for the series of conformational changes that occur in RNAs that splice by group II mechanisms. In effect, these changes have relieved the substrate pre-mRNA of the obligation to carry the sequences needed to sponsor the reaction. As the splicing apparatus has become more complex (and as the number of potential substrates has increased), proteins have played a more important role.

Reviews

260. Michel, F. and Ferat, J.-L. (1995). *Structure and activities of group II introns*. Annu. Rev. Biochem. 64, 435-461.

RNA SPLICING AND PROCESSING

5.24.12 Alternative splicing involves differential use of splice junctions

Key Terms

Alternative splicing describes the production of different RNA products from a single product by changes in the usage of splicing junctions.

Key Concepts

- Specific exons may be excluded or included in the RNA product by using or failing to use a pair of splicing junctions.
- Exons may be extended by changing one of the splice junctions to use an alternative junction.
- Sex determination in *Drosophila* involves a series of alternative splicing events in genes coding for successive products of a pathway.
- P elements of *Drosophila* show germline-specific alternative splicing.

When an interrupted gene is transcribed into an RNA that gives rise to a single type of spliced mRNA, there is no ambiguity in assignment of exons and introns. But the RNAs of some genes follow patterns of **alternative splicing**, when a single gene gives rise to more than one mRNA sequence. In some cases, the ultimate pattern of expression is dictated by the primary transcript, because the use of different startpoints or the generation of alternative 3′ ends alters the pattern of splicing. In other cases, a single primary transcript is spliced in more than one way, and internal exons are substituted, added, or deleted. In some cases, the multiple products all are made in the same cell, but in others the process is regulated so that particular splicing patterns occur only under particular conditions (for review see 246).

One of the most pressing questions in splicing is to determine what controls the use of such alternative pathways. Proteins that intervene to bias the use of alternative splice sites have been identified in two ways. In some mammalian systems, it has been possible to characterize alternative splicing *in vitro*, and to identify proteins that are required for the process. In *D. melanogaster*, aberrations in alternative splicing may be caused either by mutations in the genes that are alternatively spliced or in the genes whose products are necessary for the reaction.

Figure 24.21 shows examples of alternative splicing in which one splice site remains constant, but the other varies. The large T/ small t antigens of SV40 and the products of the adenovirus E1A region are generated by connecting a varying 5′ site to a constant 3′ site. In the case of the T/t antigens, the 5′ site used for T antigen removes a termination codon that is present in the t antigen mRNA, so that T antigen is larger than t antigen. In the case of the E1A transcripts, one of the 5′ sites connects to the last exon in a different reading frame, again making a significant change in the C-terminal part of the protein. In these examples, all the relevant

splicing events take place in every cell in which the gene is expressed, so all the protein products are made.

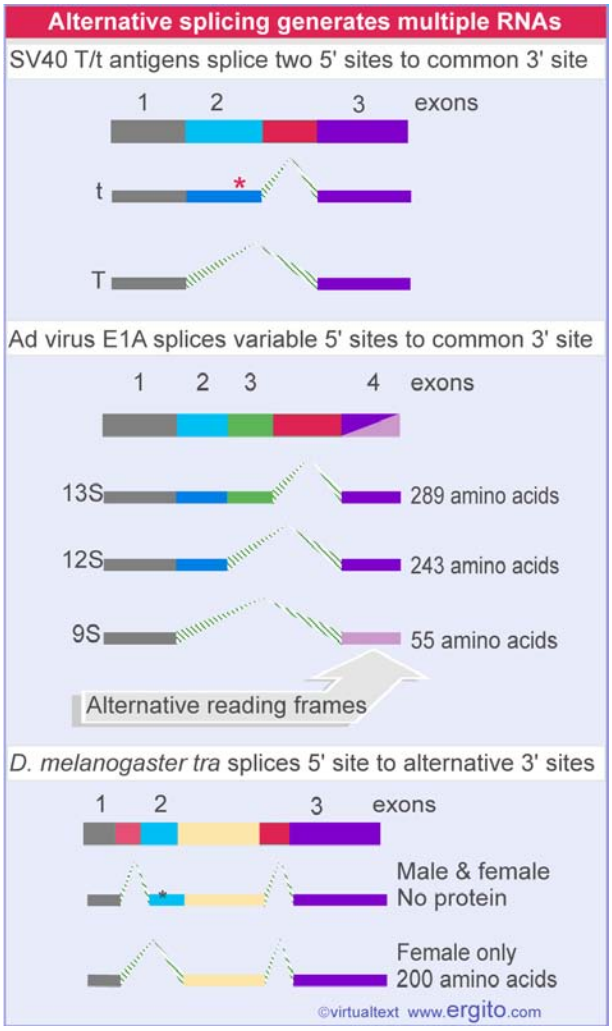

Figure 24.21 Alternative forms of splicing may generate a variety of protein products from an individual gene. Changing the splice sites may introduce termination codons (shown by asterisks) or change reading frames.

There are differences in the ratios of T/t antigens in different cell types. A protein extracted from cells that produce relatively more small t antigen can cause preferential production of small t RNA in extracts from other cell types. This protein, which was called ASF (alternative splicing factor), turns out to be the same as the splicing factor SF2, which is required for early steps in spliceosome assembly and for the first cleavage-ligation reaction (see **Figure 24.13**). ASF/SF2 is an RNA-binding protein in the SR family. When a pre-mRNA has more than one 5′ splice site preceding a single 3′ splice site, increased concentrations of ASF/SF2 promote use of the 5′ site nearest to the 3′ site at the expense of the other site (3316; 3317). This effect of ASF/SF2 can be counteracted by another splicing factor, SF5.

The exact molecular roles of the factors in controlling splice utilization are not yet known, but we see in general terms that alternative splicing involving different 5′ sites may be influenced by proteins involved in spliceosome assembly. In the case of T/t antigens, the effect probably rests on increased binding of the SR proteins to the site that is preferentially used. Alternative splicing also may be influenced by repression of one site. Exons 2 and 3 of the mouse troponin T gene are mutually exclusive; exon 2 is used in smooth muscle, but exon 3 is used in other tissues. Smooth muscle contains proteins that bind to repeated elements located on either side of exon 3, and which prevent use of the 3′ and 5′ sites that are needed to include it.

The pathway of sex determination in *D. melanogaster* involves interactions between a series of genes in which alternative splicing events distinguish male and female. The pathway takes the form illustrated in **Figure 24.22**, in which the ratio of X chromosomes to autosomes determines the expression of *sxl*, and changes in expression are passed sequentially through the other genes to *dsx*, the last in the pathway.

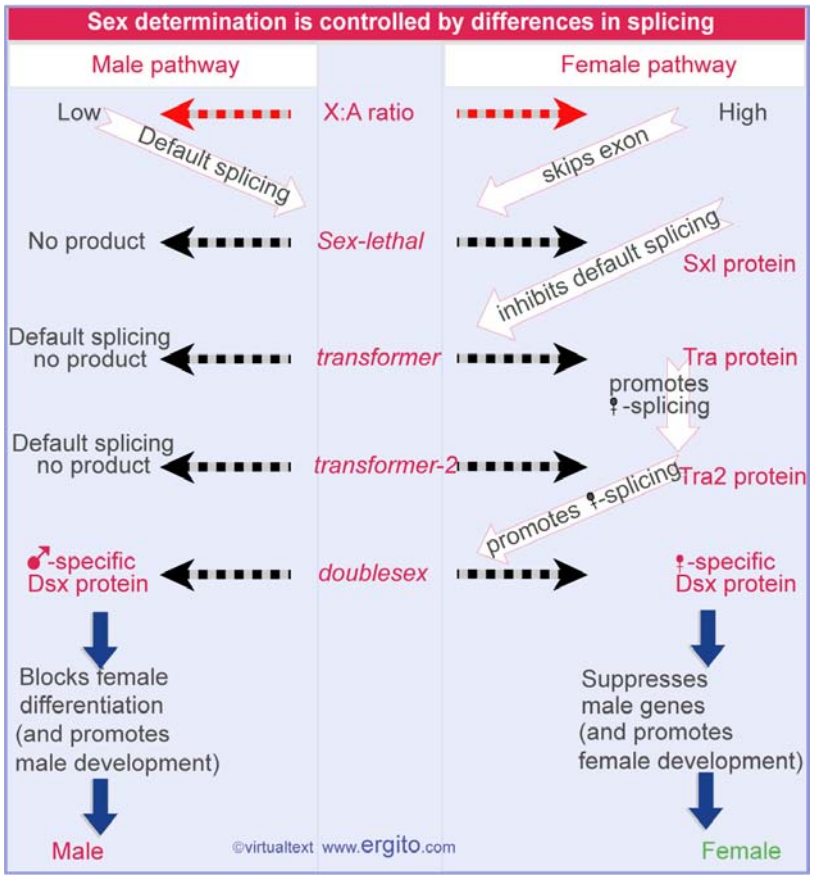

Figure 24.22 Sex determination in *D. melanogaster* involves a pathway in which different splicing events occur in females. Blocks at any stage of the pathway result in male development.

The pathway starts with sex-specific splicing of *sxl*. Exon 3 of the *sxl* gene contains a termination codon that prevents synthesis of functional protein. This exon is included in the mRNA produced in males, but is skipped in females. (Exon skipping

illustrated for another example in **Figure 24.23**.) As a result, only females produce Sxl protein. The protein has a concentration of basic amino acids that resembles other RNA-binding proteins.

The presence of Sxl protein changes the splicing of the *transformer (tra)* gene. **Figure 24.21** shows that this involves splicing a constant 5′ site to alternative 3′ sites. One splicing pattern occurs in both males and females, and results in an RNA that has an early termination codon. The presence of Sxl protein inhibits usage of the normal 3′ splice site by binding to the polypyrimidine tract at its branch site (3319). When this site is skipped, the next 3′ site is used. This generates a female-specific mRNA that codes for a protein.

So *tra* produces a protein only in females; this protein is a splicing regulator. *tra2* has a similar function in females (but is also expressed in the male germline). The Tra and Tra2 proteins are SR splicing factors that act directly upon the target transcripts. Tra and Tra2 cooperate (in females) to affect the splicing of *dsx*.

Figure 24.23 shows examples of cases in which splice sites are used to add or to substitute exons or introns, again with the consequence that different protein products are generated. In the *doublesex (dsx)* gene, females splice the 5′ site of intron 3 to the 3′ site of that intron; as a result translation terminates at the end of exon 4. Males splice the 5′ site of intron 3 directly to the 3′ site of intron 4, thus omitting exon 4 from the mRNA, and allowing translation to continue through exon 6. The result of the alternative splicing is that different proteins are produced in each sex: the male product blocks female sexual differentiation, while the female product represses expression of male-specific genes.

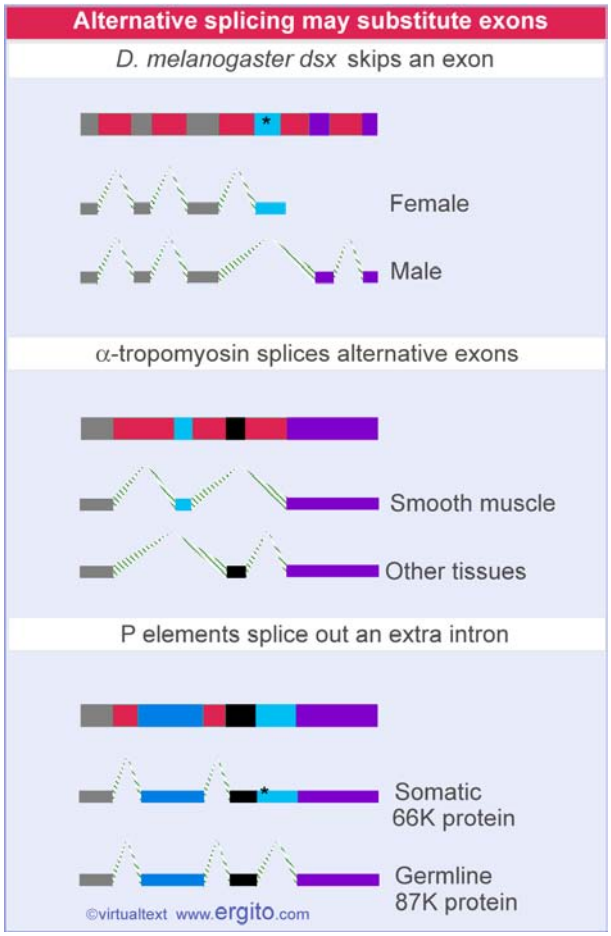

Figure 24.23 Alternative splicing events that involve both sites may cause exons to be added or substituted.

Alternative splicing of *dsx* RNA is controlled by competition between 3′ splice sites. *dsx* RNA has an element downstream of the leftmost 3′ splice site that is bound by Tra2; Tra and SR proteins associate with Tra2 at the site, which becomes an enhancer that assists binding of U2AF at the adjacent pyrimidine tract (3320; 3321). This commits the formation of the spliceosome to use this 3′ site in females rather than the alternative 3′ site. The proteins recognize the enhancer cooperatively, possibly relying on formation of some secondary structure as well as sequence *per se*.

Sex determination therefore has a pleasing symmetry: the pathway starts with a female-specific splicing event that causes omission of an exon that has a termination codon, and ends with a female-specific splicing event that causes inclusion of an exon that has a termination codon. The events have different molecular bases. At the first control point, Sxl inhibits the default splicing pattern. At the last control point, Tra and Tra2 cooperate to promote the female-specific splice.

The Tra and Tra2 proteins are not needed for normal splicing, because in their absence flies develop normally (as males). As specific regulators, they need not necessarily participate in the mechanics of the splicing reaction; in this respect they differ from SF2, which is a factor required for general splicing, but can also

influence choice of alternative splice sites.

P elements of *D. melanogaster* show a tissue-specific splicing pattern. In somatic cells, there are two splicing events, but in germline an additional splicing event removes another intron. Because a termination codon lies in the germline-specific intron, a longer protein (with different properties) is produced in germline. We discuss the consequences for control of transposition in *Molecular Biology 4.16.15 P elements are activated in the germline*, and note for now that the tissue specificity results from differences in the splicing apparatus.

The default splicing pathway of the P element pre-mRNA when the RNA is subjected to a heterologous (human) splicing extract is the germline pattern, in which intron 3 is excised. But extracts of somatic cells of *D. melanogaster* contain a protein that inhibits excision of this intron. The protein binds to sequences in exon 3; if these sequences are deleted, the intron is excised. The function of the protein is therefore probably to repress association of the spliceosome with the 5′ site of intron 3.

Reviews

246. Green, M. R. (1991). *Biochemical mechanisms of constitutive and regulated pre-mRNA splicing*. Annu. Rev. Cell Biol. 7, 559-599.

References

3316. Wu, J. Y. and Maniatis, T. (1993). *Specific interactions between proteins implicated in splice site selection and regulated alternative splicing.* Cell 75, 1061-1070.

3317. Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R., and Rottman, F. M. (1993). *General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer.* Genes Dev. 7, 2598-2608.

3319. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). *Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein.* Nature 398, 579-585.

3320. Tian, M. and Maniatis, T. (1993). *A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA.* Cell 74, 105-114.

3321. Lynch, K. W. and Maniatis, T. (1996). *Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer.* Genes Dev. 10, 2089-2101.

RNA SPLICING AND PROCESSING

5.24.13 *trans*-splicing reactions use small RNAs

Key Terms

SL RNA (Spliced leader RNA) is a small RNA that donates an exon in the *trans*-splicing reaction of trypanosomes and nematodes.

Key Concepts

- Splicing reactions usually occur only in *cis* between splice junctions on the same molecule of RNA.
- *trans*-splicing occurs in trypanosomes and worms where a short sequence (SL RNA) is spliced to the 5′ ends of many precursor mRNAs.
- SL RNA has a structure resembling the Sm-binding site of U snRNAs and may play an analogous role in the reaction.

In both mechanistic and evolutionary terms, splicing has been viewed as an *intramolecular* reaction, amounting essentially to a controlled deletion of the intron sequences at the level of RNA. In genetic terms, splicing occurs only in *cis.* This means that *only sequences on the same molecule of RNA can be spliced together.* The upper part of **Figure 24.24** shows the normal situation. The introns can be removed from each RNA molecule, allowing the exons of that RNA molecule to be spliced together, but there is no *intermolecular* splicing of exons between different RNA molecules. We cannot say that *trans* splicing never occurs between pre-mRNA transcripts of the same gene, but we know that it must be exceedingly rare, because if it were prevalent the exons of a gene would be able to complement one another genetically instead of belonging to a single complementation group.

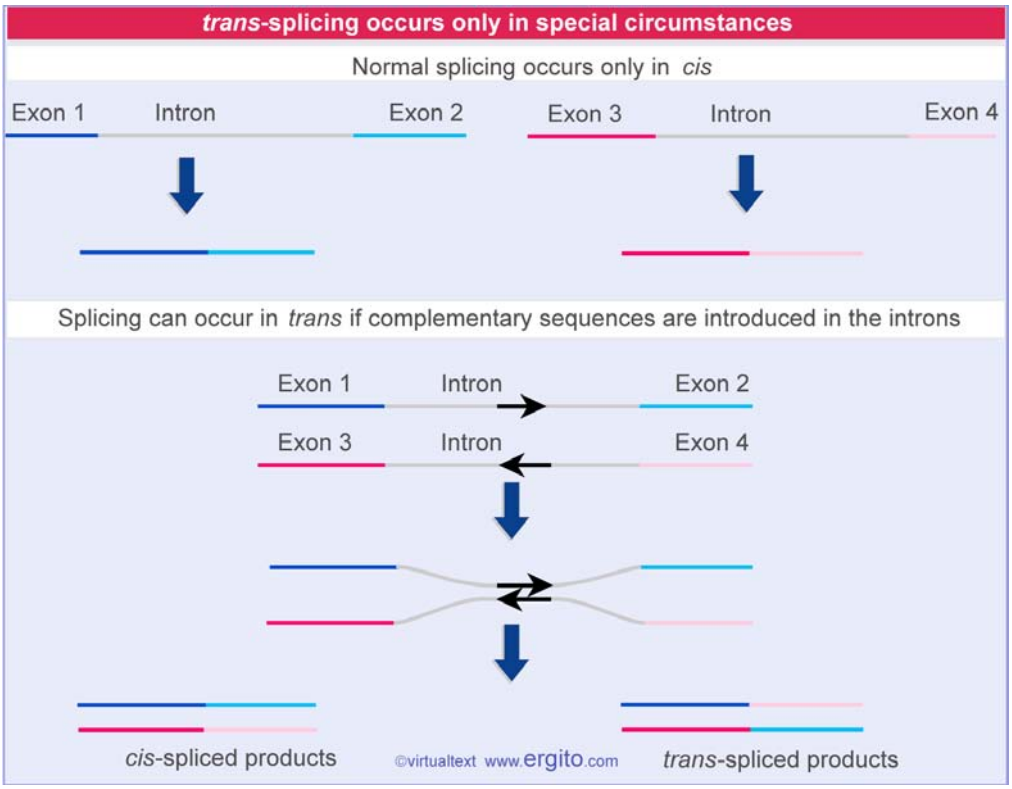

Figure 24.24 Splicing usually occurs only in *cis* between exons carried on the same physical RNA molecule, but *trans* splicing can occur when special constructs are made that support base pairing between introns.

Some manipulations can generate *trans*-splicing. In the example illustrated in the lower part of **Figure 24.24**, complementary sequences were introduced into the introns of two RNAs. Base pairing between the complements should create an H-shaped molecule. This molecule could be spliced in *cis*, to connect exons that are covalently connected by an intron, or it could be spliced in *trans*, to connect exons of the juxtaposed RNA molecules. Both reactions occur *in vitro*.

Another situation in which *trans*-splicing is possible *in vitro* occurs when substrate RNAs are provided in the form of one containing a 5′ splice site and the other containing a 3′ splice site together with appropriate downstream sequences (which may be either the next 5′ splice site or a splicing enhancer). In effect, this mimics splicing by exon definition (see the right side of **Figure 24.12**), and shows that *in vitro* it is not necessary for the left and right splice sites to be on the same RNA molecule.

These results show that there is no *mechanistic* impediment to *trans*-splicing. They exclude models for splicing that require processive movement of a spliceosome along the RNA. It must be possible for a spliceosome to recognize the 5′ and 3′ splice sites of different RNAs when they are in close proximity.

Although *trans*-splicing is rare, it occurs *in vivo* in some special situations. One is revealed by the presence of a common 35 base leader sequence at the end of numerous mRNAs in the trypanosome. But the leader sequence is not coded

upstream of the individual transcription units. Instead it is transcribed into an independent RNA, carrying additional sequences at its 3′ end, from a repetitive unit located elsewhere in the genome. **Figure 24.25** shows that this RNA carries the 35 base leader sequence followed by a 5′ splice site sequence. The sequences coding for the mRNAs carry a 3′ splice site just preceding the sequence found in the mature mRNA (730).

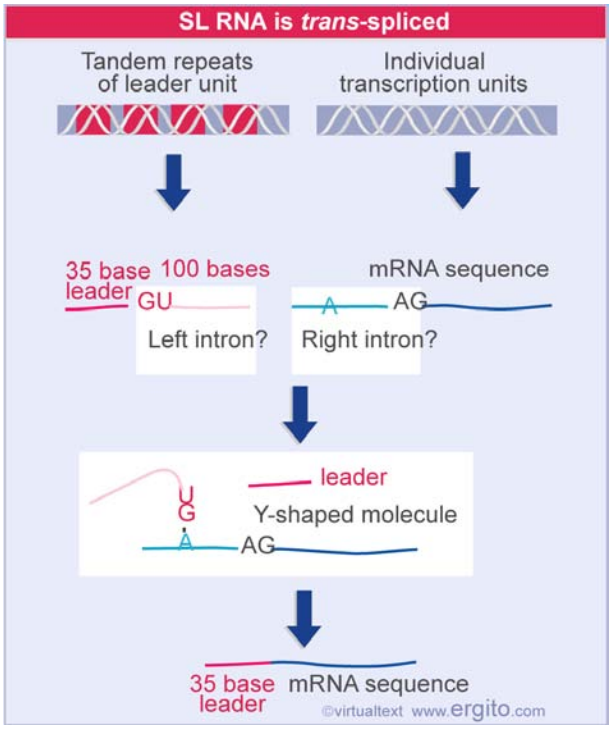

Figure 24.25 The SL RNA provides an exon that is connected to the first exon of an mRNA by *trans*-splicing. The reaction involves the same interactions as nuclear *cis*-splicing, but generates a Y-shaped RNA instead of a lariat.

When the leader and the mRNA are connected by a *trans*-splicing reaction, the 3′ region of the leader RNA and the 5′ region of the mRNA in effect comprise the 5′ and 3′ halves of an intron. When splicing occurs, a 5′–2′ link forms by the usual reaction between the GU of the 5′ intron and the branch sequence near the AG of the 3′ intron. Because the two parts of the intron are not covalently linked, this generates a Y-shaped molecule instead of a lariat (729).

A similar situation is presented by the expression of actin genes in *C. elegans.* Three actin mRNAs (and some other RNAs) share the same 22 base leader sequence at the 5′ terminus. The leader sequence is not coded in the actin gene, but is transcribed independently as part of a 100 base RNA coded by a gene elsewhere. *trans*-splicing also occurs in chloroplasts (731).

The RNA that donates the 5′ exon for *trans* splicing is called the **SL RNA** (spliced leader RNA). The SL RNAs found in several species of trypanosomes and also in the nematode (*C. elegans*) have some common features. They fold into a common secondary structure that has three stem-loops and a single-stranded region that

resembles the Sm-binding site. The SL RNAs therefore exist as snRNPs that count as members of the Sm snRNP class. Trypanosomes possess the U2, U4, and U6 snRNAs, but do not have U1 or U5 snRNAs. The absence of U1 snRNA can be explained by the properties of the SL RNA, which can carry out the functions that U1 snRNA usually performs at the 5′ splice site; thus SL RNA in effect consists of an snRNA sequence possessing U1 function, linked to the exon-intron site that it recognizes.

There are two types of SL RNA in *C. elegans.* SL1 RNA (the first to be discovered) is used for splicing to coding sequences that are preceded only by 5′ nontranslated regions (the most common situation). SL2 RNA is used in cases in which a pre-mRNA contains two coding sequences; it is spliced to the second sequence, thus releasing it from the first, and allowing it to be used as an independent mRNA (732; 733; for review see 250). About 15% of all genes in *C. elegans* are organized in transcription units that include more than one gene (most often 2-3 genes) (2862). The significance of this form of organization for control of gene expression is not clear. These transcription units do not generally resemble operons where the genes function coordinately in a pathway.

The *trans*-splicing reaction of the SL RNA may represent a step towards the evolution of the pre-mRNA splicing apparatus. The SL RNA provides in *cis* the ability to recognize the 5′ splice site, and this probably depends upon the specific conformation of the RNA. The remaining functions required for splicing are provided by independent snRNPs. The SL RNA can function without participation of proteins like those in U1 snRNP, which suggests that the recognition of the 5′ splice site depends directly on RNA.

Last updated on 8-22-2002

Reviews

250. Nilsen, T. (1993). *trans-splicing of nematode pre-mRNA*. Annu. Rev. Immunol. 47, 413-440.

References

729. Murphy, W. J., Watkins, K. P., and Agabian, N. (1986). *Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans-splicing*. Cell 47, 517-525.

730. Sutton, R. and Boothroyd, J. C. (1986). *Evidence for trans-splicing in trypanosomes*. Cell 47, 527-535.

731. Krause, M. and Hirsh, D. (1987). *A trans-spliced leader sequence on actin mRNA in C. elegans*. Cell 49, 753-761.

732. Huang, X. Y. and Hirsh, D. (1989). *A second trans-spliced RNA leader sequence in the nematode C. elegans.* Proc. Natl. Acad. Sci. USA 86, 8640-8644.

733. Hannon, G. J. et al. (1990). *trans-splicing of nematode pre-mRNA in vitro.* Cell 61, 1247-1255.

2862. Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M., and Kim, S. K. (2002). *A global analysis of C. elegans operons.* Nature 417, 851-854.

RNA SPLICING AND PROCESSING

5.24.14 Yeast tRNA splicing involves cutting and rejoining

Key Terms

An **RNA ligase** is an enzyme that functions in tRNA splicing to make a phosphodiester bond between the two exon sequences that are generated by cleavage of the intron.

Key Concepts

- tRNA splicing occurs by successive cleavage and ligation reactions.

Most splicing reactions depend on short consensus sequences and occur by transesterification reactions in which breaking and making of bonds is coordinated. The splicing of tRNA genes is achieved by a different mechanism that relies upon separate cleavage and ligation reactions.

Some 59 of the 272 nuclear tRNA genes in the yeast *S. cerevisiae* are interrupted. Each has a single intron, located just one nucleotide beyond the 3′ side of the anticodon. The introns vary in length from 14-60 bp. Those in related tRNA genes are related in sequence, but the introns in tRNA genes representing different amino acids are unrelated. *There is no consensus sequence that could be recognized by the splicing enzymes.* This is also true of interrupted nuclear tRNA genes of plants, amphibians, and mammals.

All the introns include a sequence that is complementary to the anticodon of the tRNA. This creates an alternative conformation for the anticodon arm in which the anticodon is base paired to form an extension of the usual arm. An example is drawn in **Figure 24.26**. Only the anticodon arm is affected – the rest of the molecule retains its usual structure.

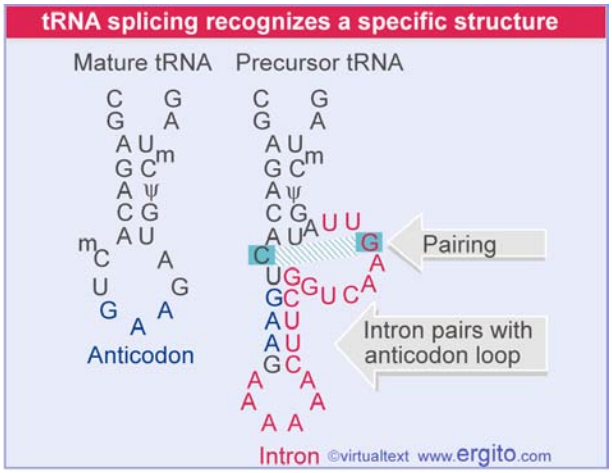

Figure 24.26 The intron in yeast $tRNA^{Phe}$ base pairs with the anticodon to change the structure of the anticodon arm. Pairing between an excluded base in the stem and the intron loop in the precursor may be required for splicing.

The exact sequence and size of the intron is not important. Most mutations in the intron do not prevent splicing. *Splicing of tRNA depends principally on recognition of a common secondary structure in tRNA rather than a common sequence of the intron.* Regions in various parts of the molecule are important, including the stretch between the acceptor arm and D arm, in the T ψ C arm, and especially the anticodon arm. This is reminiscent of the structural demands placed on tRNA for protein synthesis (see *Molecular Biology 2.6 Protein synthesis*).

The intron is not entirely irrelevant, however. Pairing between a base in the intron loop and an unpaired base in the stem is required for splicing. Mutations at other positions that influence this pairing (for example, to generate alternative patterns for pairing) influence splicing. The rules that govern availability of tRNA precursors for splicing resemble the rules that govern recognition by aminoacyl-tRNA synthetases (see *Molecular Biology 2.7.9 tRNAs are charged with amino acids by synthetases*).

In a temperature-sensitive mutant of yeast that fails to remove the introns, the interrupted precursors accumulate in the nucleus. The precursors can be used as substrates for a cell-free system extracted from wild-type cells. The splicing of the precursor can be followed by virtue of the resulting size reduction. This is seen by the change in position of the band on gel electrophoresis, as illustrated in **Figure 24.27**. The reduction in size can be accounted for by the appearance of a band representing the intron.

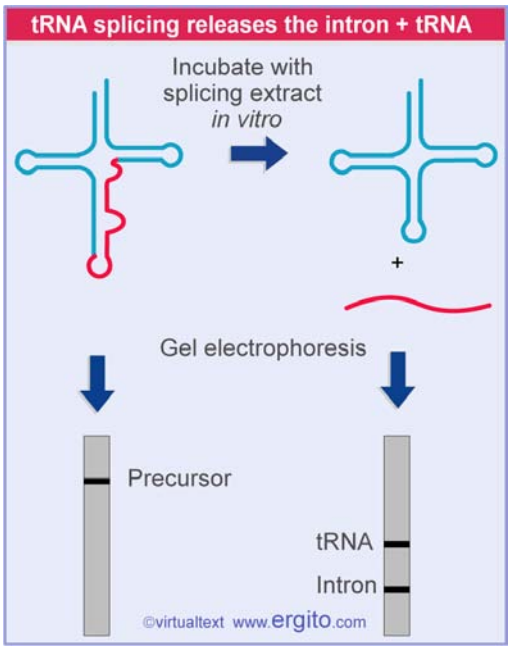

Figure 24.27 Splicing of yeast tRNA *in vitro* can be followed by assaying the RNA precursor and products by gel electrophoresis.

The cell-free extract can be fractionated by assaying the ability to splice the tRNA. The *in vitro* reaction requires ATP. Characterizing the reactions that occur with and without ATP shows that the *two separate stages of the reaction are catalyzed by different enzymes.*

- The first step does not require ATP. It involves phosphodiester bond cleavage by an atypical nuclease reaction. It is catalyzed by an endonuclease.
- The second step requires ATP and involves bond formation; it is a ligation reaction, and the responsible enzyme activity is described as an **RNA ligase**.

Last updated on 8-29-2002

RNA SPLICING AND PROCESSING

5.24.15 The splicing endonuclease recognizes tRNA

Key Concepts

- An endonuclease cleaves the tRNA precursors at both ends of the intron.
- The yeast endonuclease is a heterotetramer, with two (related) catalytic subunits.
- It uses a measuring mechanism to determine the sites of cleavage by their positions relative to a point in the tRNA structure.
- The archaeal nuclease has a simpler structure and recognizes a bulge-helix-bulge structural motif in the substrate.

The endonuclease is responsible for the specificity of intron recognition. It cleaves the precursor at both ends of the intron. The yeast endonuclease is a heterotetrameric protein. Its activities are illustrated in **Figure 24.28**. The related subunits Sen34 and Sen2 cleave the 3 ′ and 5 ′ splice sites, respectively. Subunit Sen54 may determine the sites of cleavage by "measuring" distance from a point in the tRNA structure. This point is in the elbow of the (mature) L-shaped structure (2878). The role of subunit Sen15 is not known, but its gene is essential in yeast. The base pair that forms between the first base in the anticodon loop and the base preceding the 3 ′ splice site is required for 3 ′ splice site cleavage (735; 736; 737; 2880).

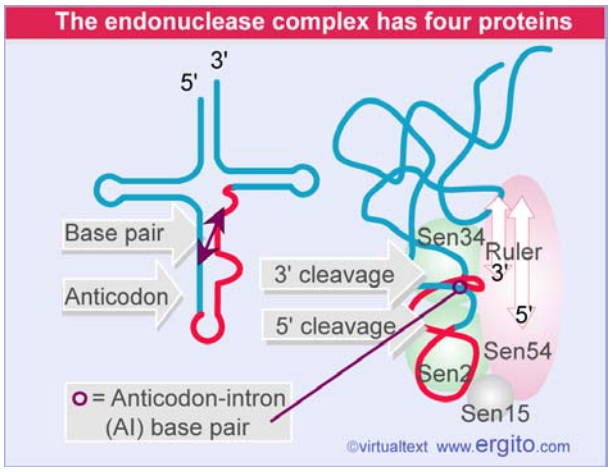

Figure 24.28 The 3 ′ and 5 ′ cleavages in *S. cerevisiae* pre-tRNA are catalyzed by different subunits of the endonuclease. Another subunit may determine location of the cleavage sites by measuring distance from the mature structure. The AI base pair is also important.

An interesting insight into the evolution of tRNA splicing is provided by the endonucleases of archaea. These are homodimers or homotetramers, in which each subunit has an active site (although only two of the sites function in the tetramer) that

cleaves one of the splice sites (2879). The subunit has sequences related to the sequences of the active sites in the Sen34 and Sen2 subunits of the yeast enzyme. However, the archaeal enzymes recognize their substrates in a different way. Instead of measuring distance from particular sequences, they recognize a structural feature, called the bulge-helix-bulge. **Figure 24.29** shows that cleavage occurs in the two bulges (2877; 2876).

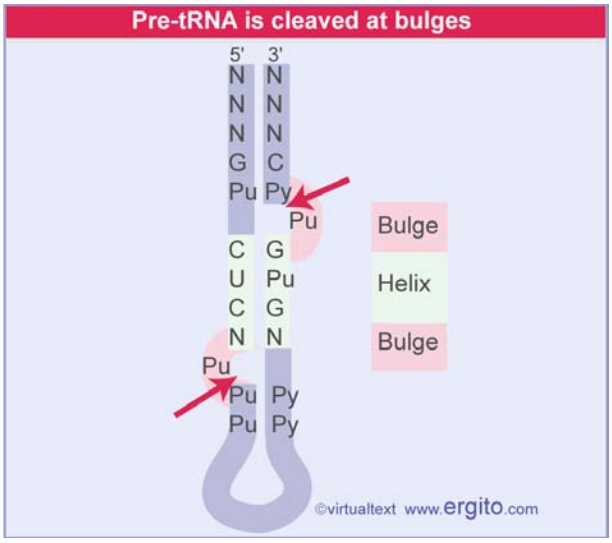

Figure 24.29 Archaeal tRNA splicing endonuclease cleaves each strand at a bulge in a bulge-helix-bulge motif.

So the origin of splicing of tRNA precedes the separation of the archaea and the eukaryotes. If it originated by insertion of the intron into tRNAs, this must have been a very ancient event.

Last updated on 8-29-2002

References

735. Reyes, V. M. and Abelson, J. (1988). *Substrate recognition and splice site determination in yeast tRNA splicing*. Cell 55, 719-730.

736. Mattoccia, E. et al. (1988). *Site selection by the tRNA splicing endonuclease of X. laevis*. Cell 55, 731-738.

737. Baldi, I. M. et al. (1992). *Participation of the intron in the reaction catalyzed by the Xenopus tRNA splicing endonuclease*. Science 255, 1404-1408.

2876. Diener, J. L. and Moore, P. B. (1998). *Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif.* Mol. Cell 1, 883-894.

2877. Lykke-Andersen, J. and Garrett, R. A. (1997). *RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history.* EMBO J. 16, 6290-6300.

2878. Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997). *The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases.* Cell 89, 849-858.

2879. Kleman-Leyer, K., Armbruster, D. W., and Daniels, C. J. (2000). *Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems.* Cell 89, 839-847.

2880. Di Nicola Negri, E., Fabbri, S., Bufardeci, E., Baldi, M. I., Mattoccia, E., and Tocchini-Valentini, G. P. (1997). *The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements.* Cell 89, 859-866.

RNA SPLICING AND PROCESSING

5.24.16 tRNA cleavage and ligation are separate reactions

Key Concepts

- Release of the intron generates two half-tRNAs that pair to form the mature structure.
- The halves have the unusual ends 5′ hydroxyl and 2′–3′ cyclic phosphate.
- The 5′–OH end is phosphorylated by a polynucleotide kinase, the cyclic phosphate group is opened by phosphodiesterase to generate a 2′–phosphate terminus and 3′–OH group, exon ends are joined by an RNA ligase, and the 2′–phosphate is removed by a phosphatase.

The overall tRNA splicing reaction is summarized in **Figure 24.30**. The products of cleavage are a linear intron and two half-tRNA molecules. These intermediates have unique ends. Each 5′ terminus ends in a hydroxyl group; each 3′ terminus ends in a 2′,3′–cyclic phosphate group. (All other known RNA splicing enzymes cleave on the other side of the phosphate bond.)

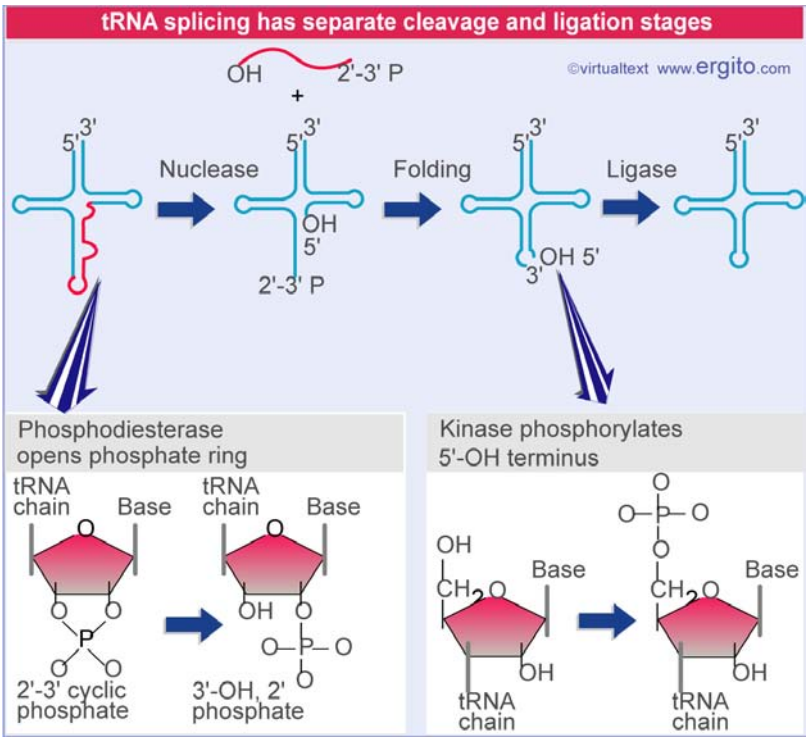

Figure 24.30 Splicing of tRNA requires separate nuclease and ligase activities. The exon-intron boundaries are cleaved by the nuclease to generate 2′-3′ cyclic phosphate and 5′ OH termini. The cyclic phosphate is opened to generate 3′-OH and 2′ phosphate groups. The 5′-OH is phosphorylated. After releasing the intron, the tRNA half molecules fold into a tRNA-like structure that now has a 3′-OH, 5′-P break. This is sealed by a ligase.

The two half-tRNAs base pair to form a tRNA-like structure. When ATP is added, the second reaction occurs. Both of the unusual ends generated by the endonuclease must be altered.

The cyclic phosphate group is opened to generate a 2′–phosphate terminus. This reaction requires cyclic phosphodiesterase activity. The product has a 2′–phosphate group and a 3′–OH group.

The 5′–OH group generated by the nuclease must be phosphorylated to give a 5′–phosphate. This generates a site in which the 3′–OH is next to the 5′–phosphate. Covalent integrity of the polynucleotide chain is then restored by ligase activity.

All three activities – phosphodiesterase, polynucleotide kinase, and adenylate synthetase (which provides the ligase function) – are arranged in different functional domains on a single protein. They act sequentially to join the two tRNA halves.

The spliced molecule is now uninterrupted, with a 5′–3′ phosphate linkage at the site of splicing, but it also has a 2′–phosphate group marking the event. The surplus group must be removed by a phosphatase.

Generation of a 2′,3′–cyclic phosphate also occurs during the tRNA-splicing reaction in plants and mammals. The reaction in plants seems to be the same as in

yeast, but the detailed chemical reactions are different in mammals.

The yeast tRNA precursors also can be spliced in an extract obtained from the germinal vesicle (nucleus) of *Xenopus* oocytes. This shows that the reaction is not species-specific. *Xenopus* must have enzymes able to recognize the introns in the yeast tRNAs.

The ability to splice the products of tRNA genes is therefore well conserved, but is likely to have a different origin from the other splicing reactions (such as that of nuclear pre-mRNA). The tRNA-splicing reaction uses cleavage and synthesis of bonds and is determined by sequences that are external to the intron. Other splicing reactions use transesterification, in which bonds are transferred directly, and the sequences required for the reaction lie within the intron.

Last updated on 8-29-2002

RNA SPLICING AND PROCESSING

5.24.17 The unfolded protein response is related to tRNA splicing

Key Concepts

- Ire1p is an inner nuclear membrane protein with its N-terminal domain in the ER lumen, and its C-terminal domain in the nucleus.
- Binding of an unfolded protein to the N-terminal domain activates the C-terminal nuclease by autophosphorylation.
- The activated nuclease cleaves Hac1 mRNA to release an intron and generate exons that are ligated by a tRNA ligase.
- The spliced Hac1 mRNA codes for a transcription factor that activates genes coding for chaperones that help to fold unfolded proteins.

An unusual splicing system that is related to tRNA splicing mediates the response to unfolded proteins in yeast. The accumulation of unfolded proteins in the lumen of the ER triggers a response pathway that leads to increased transcription of genes coding for chaperones that assist protein folding in the ER. A signal must therefore be transmitted from the lumen of the ER to the nucleus.

The sensor that activates the pathway is the protein Ire1p. It is an integral membrane protein (Ser/Thr) kinase that has domains on each side of the ER membrane. The N-terminal domain in the lumen of the ER detects the presence of unfolded proteins, presumably by binding to exposed motifs. This causes aggregation of monomers and activates the C-terminal domain on the other side of the membrane by autophosphorylation.

Genes that are activated by this pathway have a common promoter element, the UPRE (unfolded protein response element). The transcription factor Hac1p binds to the UPRE, and is produced in response to accumulation of unfolded proteins. The trigger for production of Hac1p is the action of Ire1p on Hac1 mRNA.

The operation of the pathway is summarized in **Figure 24.31**. Under normal conditions, when the pathway is not activated, Hac1 mRNA is translated into a protein that is rapidly degraded. The activation of Ire1p results in the splicing of the Hac1 mRNA to change the sequence of the protein to a more stable form. This form provides the functional transcription factor that activates genes with the UPRE.

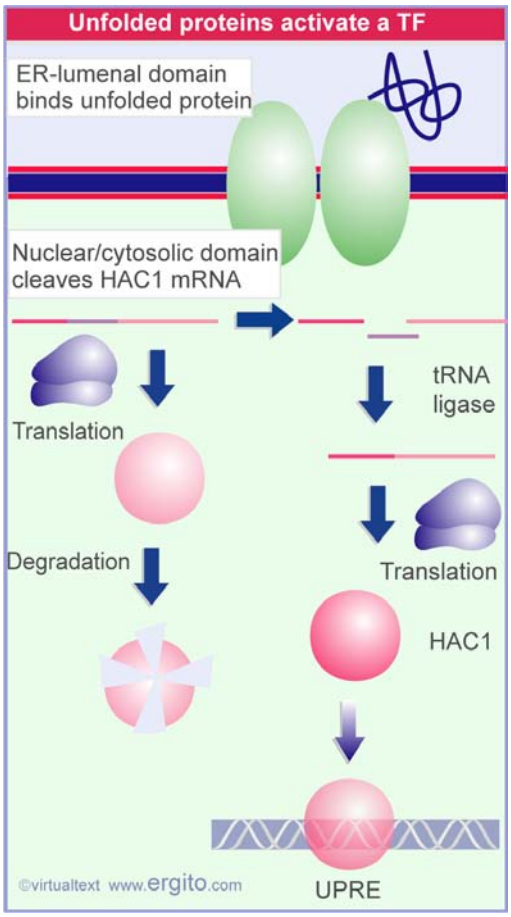

Figure 24.31 The unfolded protein response occurs by activating special splicing of HAC1 mRNA to produce a transcription factor that recognizes the UPRE.

Unusual splicing components are involved in this reaction. Ire1P has an endonuclease activity that acts directly on Hac1 mRNA to cleave the two splicing junctions (739). The two junctions are ligated by the tRNA ligase that acts in the tRNA splicing pathway (738). The endonuclease reaction resembles the cleavage of tRNA during splicing (3322).

Where does the modification of Hac1 mRNA occur? Ire1p is probably located in the inner nuclear membrane, with the N-terminal sensor domain in the ER lumen, and the C-terminal kinase/nuclease domain in the nucleus. This would it enable it to act directly on Hac1 RNA before it is exported to the cytoplasm. It also would allow easy access by the tRNA ligase. There is no apparent relationship between the Ire1p nuclease activity and the tRNA splicing endonuclease, so it is not obvious how this specialized system would have evolved.

References

738. Sidrauski, C., Cox, J. S., and Walter, P. (1996). *tRNA ligase is required for regulated mRNA splicing in the unfolded protein response*. Cell 87, 405-413.

739. Sidrauski, C. and Walter, P. (1997). *The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response*. Cell 90, 1031-1039.

3322. Gonzalez, T. N., Sidrauski, C., Dorfler, S., and Walter, P. (1999). *Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway*. EMBO J. 18, 3119-3132.

RNA SPLICING AND PROCESSING

5.24.18 The 3′ ends of polI and polIII transcripts are generated by termination

Key Concepts

- RNA polymerase I terminates transcription at an 18 base terminator sequence.
- RNA polymerase III terminates transcription in poly(U)$_4$ sequence embedded in a G·C-rich sequence.

3′ ends of RNAs can be generated in two ways. Some RNA polymerases terminate transcription at a defined (terminator) sequence in DNA, as shown in **Figure 24.32**. Other RNA polymerases do not show discrete termination, but continue past the site corresponding to the 3′ end, which is generated by cleavage of the RNA by an endonuclease, as shown in **Figure 24.33**.

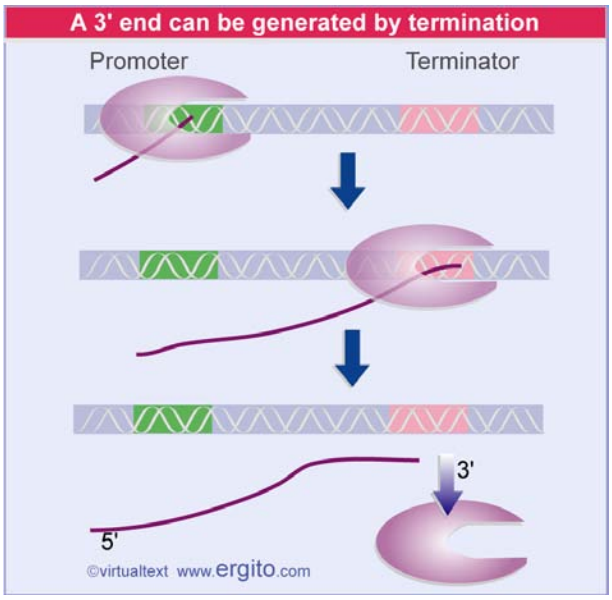

Figure 24.32 When a 3′ end is generated by termination, RNA polymerase and RNA are released at a discrete (terminator) sequence in DNA.

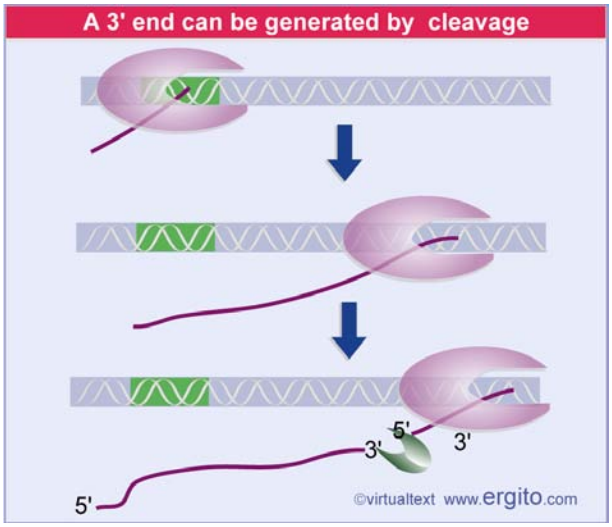

Figure 24.33 When a 3′ end is generated by cleavage, RNA polymerase continues transcription while an endonuclease cleaves at a defined sequence in the RNA.

Information about the termination reaction for eukaryotic RNA polymerases is less detailed than our knowledge of initiation. RNA polymerases I and III have discrete termination events (like bacterial RNA polymerase), but it is not clear whether RNA polymerase II usually terminates in this way.

For RNA polymerase I, the sole product of transcription is a large precursor that contains the sequences of the major rRNA. The precursor is subjected to extensive processing. Termination occurs at a discrete site >1000 bp downstream of the mature 3′ end, which is generated by cleavage. Termination involves recognition of an 18 base terminator sequence by an ancillary factor.

With RNA polymerase III, transcription *in vitro* generates molecules with the same 5′ and 3′ ends as those synthesized *in vivo*. The termination reaction resembles intrinsic termination by bacterial RNA polymerase (see *Molecular Biology 3.9.21 There are two types of terminators in E. coli*). Termination usually occurs at the second U within a run of 4 U bases, but there is heterogeneity, with some molecules ending in 3 or even 4 U bases. The same heterogeneity is seen in molecules synthesized *in vivo*, so it seems to be a *bona fide* feature of the termination reaction.

Just like the prokaryotic terminators, the U run is embedded in a G·C-rich region. Although sequences of dyad symmetry are present, they are not needed for termination, since mutations that abolish the symmetry do not prevent the normal completion of RNA synthesis. Nor are any sequences beyond the U run necessary, since all distal sequences can be replaced without any effect on termination.

The U run itself is not sufficient for termination, because regions of 4 successive U residues exist within transcription units read by RNA polymerase III. (However, there are no internal U_5 runs, which fits with the greater efficiency of termination when the terminator is a U_5 rather than U_4 sequence.) The critical feature in termination must therefore be the recognition of a U_4 sequence in a context that is

rich in G·C base pairs.

How does the termination reaction occur? It cannot rely on the weakness of the rU-dA RNA-DNA hybrid region that lies at the end of the transcript, because often only the first two U residues are transcribed. Perhaps the G·C-rich region plays a role in slowing down the enzyme, but there does not seem to be a counterpart to the hairpin involved in prokaryotic termination. We remain puzzled how the enzyme can respond so specifically to such a short signal. And in contrast with the initiation reaction, which RNA polymerase III cannot accomplish alone, termination seems to be a function of the enzyme itself.

RNA SPLICING AND PROCESSING

5.24.19 The 3′ ends of mRNAs are generated by cleavage and polyadenylation

Key Terms

Cordycepin is 3′ deoxyadenosine, an inhibitor of polyadenylation of RNA.

Endonucleases cleave bonds within a nucleic acid chain; they may be specific for RNA or for single-stranded or double-stranded DNA.

Poly(A) polymerase is the enzyme that adds the stretch of polyadenylic acid to the 3′ of eukaryotic mRNA. It does not use a template.

Key Concepts

- The sequence AAUAAA is a signal for cleavage to generate a 3′ end of mRNA that is polyadenylated.
- The reaction requires a protein complex that contains a specificity factor, an endonuclease, and poly(A) polymerase.
- The specificity factor and endonuclease cleave RNA downstream of AAUAAA.
- The specificity factor and poly(A) polymerase add ~200 A residues processively to the 3′ end.
- A·U-rich sequences in the 3′ tail control cytoplasmic polyadenylation or deadenylation during *Xenopus* embryonic development.

It is not clear whether RNA polymerase II actually engages in a termination event at a specific site. It is possible that its termination is only loosely specified. In some transcription units, termination occurs >1000 bp downstream of the site corresponding to the mature 3′ end of the mRNA (which is generated by cleavage at a specific sequence). Instead of using specific terminator sequences, the enzyme ceases RNA synthesis within multiple sites located in rather long "terminator regions." The nature of the individual termination sites is not known.

The 3′ ends of mRNAs are generated by cleavage followed by polyadenylation. Addition of poly(A) to nuclear RNA can be prevented by the analog 3′–deoxyadenosine, also known as **cordycepin**. Although cordycepin does not stop the transcription of nuclear RNA, its addition prevents the appearance of mRNA in the cytoplasm. This shows that polyadenylation is *necessary* for the maturation of mRNA from nuclear RNA.

Generation of the 3′ end is illustrated in **Figure 24.34**. RNA polymerase transcribes past the site corresponding to the 3′ end, and sequences in the RNA are recognized as targets for an endonucleolytic cut followed by polyadenylation. A single processing complex undertakes both the cutting and polyadenylation. The polyadenylation stabilizes the mRNA against degradation from the 3′ end. Its 5′ end

is already stabilized by the cap. RNA polymerase continues transcription after the cleavage, but the 5′ end that is generated by the cleavage is unprotected. As a result, the rest of the transcript is rapidly degraded. This makes it difficult to determine what is happening beyond the point of cleavage.

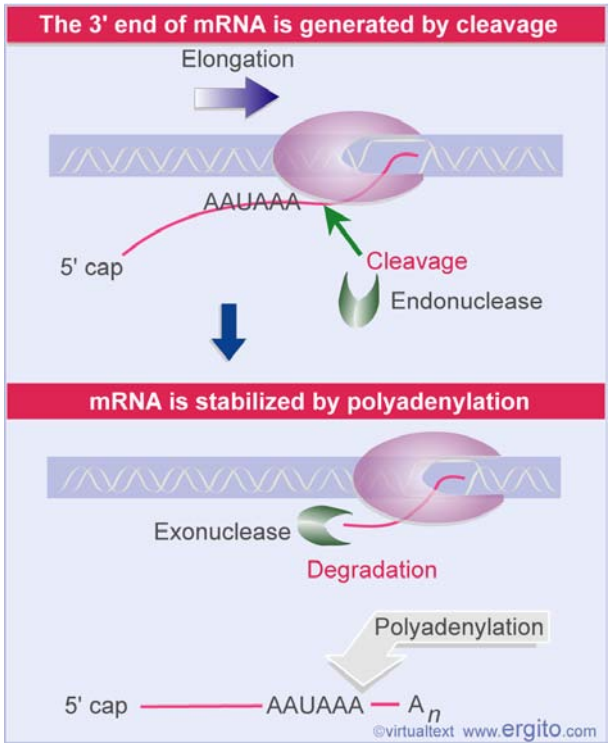

Figure 24.34 The sequence AAUAAA is necessary for cleavage to generate a 3′ end for polyadenylation.

A common feature of mRNAs in higher eukaryotes (but not in yeast) is the presence of the highly conserved sequence AAUAAA in the region from 11-30 nucleotides upstream of the site of poly(A) addition. Deletion or mutation of the AAUAAA hexamer prevents generation of the polyadenylated 3′ end. The signal is needed for both cleavage and polyadenylation (744; 745; for review see 248).

The development of a system in which polyadenylation occurs *in vitro* opened the route to analyzing the reactions. The formation and functions of the complex that undertakes 3′ processing are illustrated in **Figure 24.35**. Generation of the proper 3′ terminal structure requires an **endonuclease** (consisting of the components CFI and CFII) to cleave the RNA, a **poly(A) polymerase** (PAP) to synthesize the poly(A) tail, and a *specificity component* (CPSF) that recognizes the AAUAAA sequence and directs the other activities. A stimulatory factor, CstF, binds to a G-U-rich sequence that is downstream from the cleavage site itself (746).

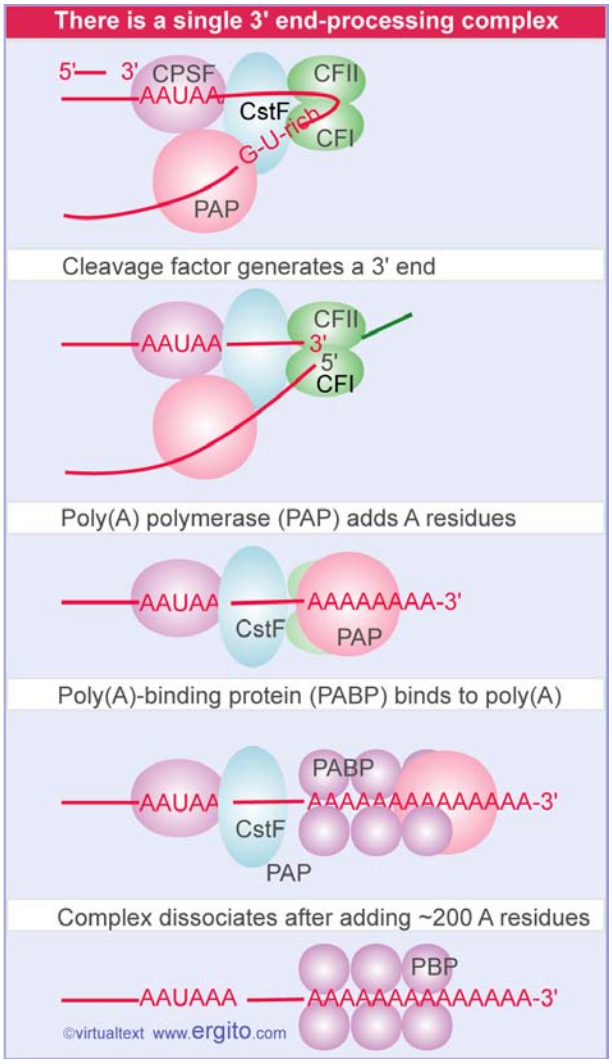

Figure 24.35 The 3 ′ processing complex consists of several activities. CPSF and CstF each consist of several subunits; the other components are monomeric. The total mass is >900 kD.

The specificity factor contains 4 subunits, which together bind specifically to RNA containing the sequence AAUAAA. The individual subunits are proteins that have common RNA-binding motifs, but which by themselves bind nonspecifically to RNA. Protein-protein interactions between the subunits may be needed to generate the specific AAUAAA-binding site. CPSF binds strongly to AAUAAA only when CstF is also present to bind to the G-U-rich site.

The specificity factor is needed for both the cleavage and polyadenylation reactions. It exists in a complex with the endonuclease and poly(A) polymerase, and this complex usually undertakes cleavage followed by polyadenylation in a tightly coupled manner.

The two components CFI and CFII (cleavage factors I and II), together with specificity factor, are necessary and sufficient for the endonucleolytic cleavage.

The poly(A) polymerase has a nonspecific catalytic activity. When it is combined with the other components, the synthetic reaction becomes specific for RNA containing the sequence AAUAAA. The polyadenylation reaction passes through two stages. First, a rather short oligo(A) sequence (~10 residues) is added to the 3′ end. This reaction is absolutely dependent on the AAUAAA sequence, and poly(A) polymerase performs it under the direction of the specificity factor. In the second phase, the oligo(A) tail is extended to the full ~200 residue length. This reaction requires another stimulatory factor that recognizes the oligo(A) tail and directs poly(A) polymerase specifically to extend the 3′ end of a poly(A) sequence.

The poly(A) polymerase by itself adds A residues individually to the 3′ position. Its intrinsic mode of action is distributive; it dissociates after each nucleotide has been added. However, in the presence of CPSF and PABP (poly(A)-binding protein), it functions processively to extend an individual poly(A) chain. The PABP is a 33 kD protein that binds stoichiometrically to the poly(A) stretch. The length of poly(A) is controlled by the PABP, which in some way limits the action of poly(A) polymerase to ~200 additions of A residues. The limit may represent the accumulation of a critical mass of PABP on the poly(A) chain. PABP binds to the translation initiation factor eIF4G, thus generating a closed loop in which a protein complex contains both the 5′ and 3′ ends of the mRNA (see **Figure 6.20** in *Molecular Biology 2.6.9 Eukaryotes use a complex of many initiation factors*).

Polyadenylation is an important determinant of mRNA function. It may affect both stability and initiation of translation (see *Molecular Biology 2.5.10 The 3′ terminus is polyadenylated*). In embryonic development in some organisms, the presence of poly(A) is used to control translation, and pre-existing mRNAs may either be polyadenylated (to stimulate translation) or deadenylated (to terminate translation). During *Xenopus* embryonic development, polyadenylation of mRNA in the cytoplasm in *Xenopus* depends on a specific *cis*-acting element (the CPE) in the 3′ tail. This is another AU-rich sequence, UUUUUAU (2313; 2314).

In *Xenopus* embryos at least two type of *cis*-acting sequences found in the 3′ tail can trigger deadenylation. EDEN (embryonic deadenylation element) is a 17 nucleotide sequence (2310). ARE elements are AU-rich, usually containing tandem repeats of AUUUA (2311). There is a poly(A)-specific RNAase (PARN) that could be involved in the degradation (2312). Of course, deadenylation is not always triggered by specific elements; in some situations (including the normal degradation of mRNA as it ages), poly(A) is degraded unless it is specifically stabilized.

Last updated on 1-22-2002

Reviews

248. Wahle, E. and Keller, W. (1992). *The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors*. Annu. Rev. Biochem. 61, 419-440.

References

744. Conway, L. and Wickens, M. (1985). *A sequence downstream of AAUAAA is required for formation of SV40 late mRNA 3′ termini in frog oocytes*. Proc. Natl. Acad. Sci. USA 82, 3949-3953.

745. Gil, A. and Proudfoot, N. (1987). *Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3′ end formation*. Cell 49, 399-406.

746. Takagaki, Y., Ryner, L. C., and Manley, J. L. (1988). *Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation*. Cell 52, 731-742.

2310. Bouvet, P., Omilli, F., Arlot-Bonnemains, Y., Legagneux, V., Roghi, C., Bassez, T., and Osborne, H. B. (1994). *The deadenylation conferred by the 3′ untranslated region of a developmentally controlled mRNA in Xenopus embryos is switched to polyadenylation by deletion of a short sequence element.* Mol. Cell Biol. 14, 1893-1900.

2311. Voeltz, G. K. and Steitz, J. A. (1998). *AUUUA sequences direct mRNA deadenylation uncoupled from decay during Xenopus early development.* Mol. Cell Biol. 18, 7537-7545.

2312. Karner, C. G., Wormington, M., Muckenthaler, M., Schneider, S., Dehlin, E., and Wahle, E. (1998). *The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of Xenopus oocytes.* EMBO J. 17, 5427-5437.

2313. Fox, C. A., Sheets, M. D., and Wickens, M. P. (1989). *Poly(A) addition during maturation of frog oocytes: distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU.* Genes Dev. 3, 2151-2162.

2314. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. (1989). *Poly(A) elongation during Xenopus oocyte maturation is required for translational recruitment and is mediated by a short sequence element.* Genes Dev. 3, 803-815.

RNA SPLICING AND PROCESSING

5.24.20 Cleavage of the 3′ end of histone mRNA may require a small RNA

Key Concepts

- Histone mRNAs are not polyadenylated; their 3′ ends are generated by a cleavage reaction that depends on the structure of the mRNA.
- The cleavage reaction requires the SLBP to bind to a stem-loop structure, and the U7 snRNA to pair with an adjacent single-stranded region.

Some mRNAs are not polyadenylated. The formation of their 3′ ends is therefore different from the coordinated cleavage/polyadenylation reaction. The most prominent members of this mRNA class are the mRNAs coding for histones that are synthesized during DNA replication. Formation of their 3′ ends depends upon secondary structure. The structure at the 3' terminus is a highly conserved stem-loop structure, with a stem of 6 bp and a loop of 4 nucleotides. Cleavage occurs 4-5 bases downstream of the stem-loop. Two factors are required for the cleavage reaction: the stem-loop binding protein (SLBP) recognizes the structure (3324); and the U7 snRNA pairs with a purine-rich sequence (the histone downstream element, or HDE) located ~10 nucleotides downstream of the cleavage site (743; 3325).

Mutations that prevent formation of the duplex stem of the stem-loop prevent formation of the end of the RNA. Secondary mutations that restore duplex structure (though not necessarily the original sequence) behave as revertants. This suggests that *formation of the secondary structure is more important than the exact sequence.* The SLBP binds to the stem-loop and then interacts with U7 snRNP to enhance its interaction with the downstream binding site for U7 snRNA (3326). U7 snRNP is a minor snRNP consisting of the 63 nucleotide U7 snRNA and a set of several proteins (including Sm proteins; see *Molecular Biology 5.24.5 snRNAs are required for splicing*).

The reaction between histone H3 mRNA and U7 snRNA is drawn in **Figure 24.36**. The upstream hairpin and the HDE that pairs with U7 snRNA are conserved in histone H3 mRNAs of several species. The U7 snRNA has sequences towards its 5′ end that pair with the histone mRNA consensus sequences. 3′ processing is inhibited by mutations in the HDE that reduce ability to pair with U7 snRNA. Compensatory mutations in U7 snRNA that restore complementarity also restore 3′ processing (3327). This suggests that U7 snRNA functions by base pairing with the histone mRNA. The sequence of the HDE varies among the various histone mRNAs, with the result that binding of snRNA is not by itself necessarily stable, but requires also the interaction with SLBP

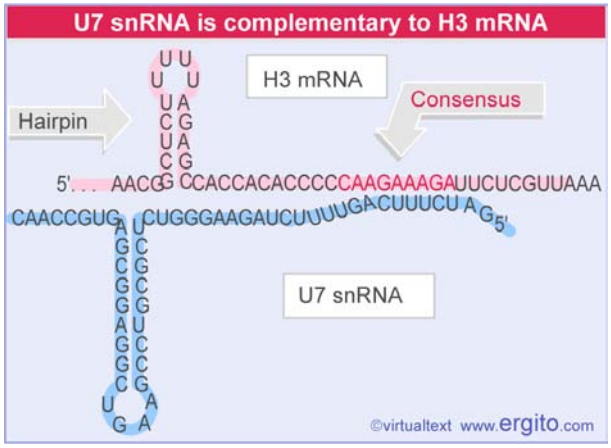

Figure 24.36 Generation of the 3′ end of histone H3 mRNA depends on a conserved hairpin and a sequence that base pairs with U7 snRNA.

Cleavage to generate a 3′ terminus occurs a fixed distance from the site recognized by U7 snRNA, which suggests that the snRNA is involved in defining the cleavage site (for review see 241). However, the factor(s) actually responsible for cleavage have not yet been identified.

Last updated on 1-20-2003

Reviews

241. Birnstiel, M. L. (1985). *Transcription termination and 3′ processing: the end is in site*. Cell 41, 349-359.

References

743. Galli, G. et al. (1983). *Biochemical complementation with RNA in the Xenopus oocyte: a small RNA is required for the generation of 3′ histone mRNA termini*. Cell 34, 823-828.

3324. Wang, Z. F., Whitfield, M. L., Ingledue, T. C., Dominski, Z., and Marzluff, W. F. (1996). *The protein that binds the 3′ end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing.* Genes Dev. 10, 3028-3040.

3325. Mowry, K. L. and Steitz, J. A. (1987). *Identification of the human U7 snRNP as one of several factors involved in the 3′ end maturation of histone premessenger RNA""s.* Science 238, 1682-1687.

3326. Dominski, Z., Erkmann, J. A., Greenland, J. A., and Marzluff, W. F. (2001). *Mutations in the RNA binding domain of stem-loop binding protein define separable requirements for RNA binding and for histone pre-mRNA processing.* Mol. Cell. Biol. 21, 2008-2017.

3327. Bond, U. M., Yario, T. A., and Steitz, J. A. (1991). *Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vitro and in vitro.* Genes Dev. 5, 1709-1722.

RNA SPLICING AND PROCESSING

5.24.21 Production of rRNA requires cleavage events

Key Terms

45S RNA is a precursor that contains the sequences of both major ribosomal RNAs (28S and 18S rRNAs).

Key Concepts

- The large and small rRNAs are released by cleavage from a common precursor RNA.

The major rRNAs are synthesized as part of a single primary transcript that is processed to generate the mature products. The precursor contains the sequences of the 18S, 5.8S, and 28S rRNAs. In higher eukaryotes, the precursor is named for its sedimentation rate as **45S RNA**. In lower eukaryotes, it is smaller (35S in yeast).

The mature rRNAs are released from the precursor by a combination of cleavage events and trimming reactions (for review see 980). **Figure 24.37** shows the general pathway in yeast. There can be variations in the order of events, but basically similar reactions are involved in all eukaryotes. Most of the 5′ ends are generated directly by a cleavage event. Most of the 3′ ends are generated by cleavage followed by a 3′–5′ trimming reaction.

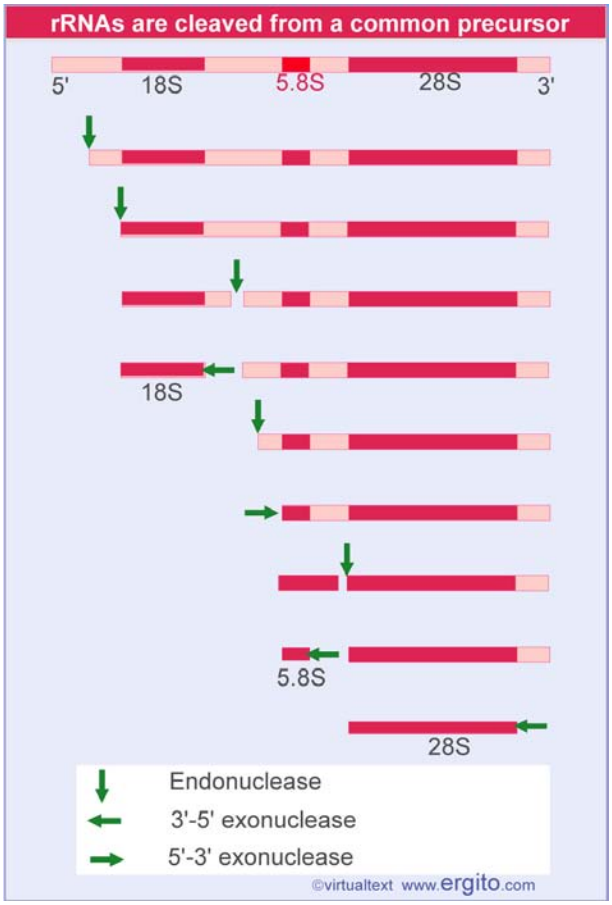

Figure 24.37 Mature eukaryotic rRNAs are generated by cleavage and trimming events from a primary transcript.

Many ribonucleases have been implicated in processing rRNA, including the exosome, an assembly of several exonucleases that also participates in mRNA degradation (see *Molecular Biology 2.5.13 mRNA degradation involves multiple activities*). Mutations in individual enzymes usually do not prevent processing, suggesting that their activities are redundant and that different combinations of cleavages can be used to generate the mature molecules.

There are always multiple copies of the transcription unit for the rRNAs. The copies are organized as tandem repeats (see *Molecular Biology 1.4.9 The repeated genes for rRNA maintain constant sequence*).

5S RNA is transcribed from separate genes by RNA polymerase III. Usually the 5S genes are clustered, but are separate from the genes for the major rRNAs. (In the case of yeast, a 5S gene is associated with each major transcription unit, but is transcribed independently.)

There is a difference in the organization of the precursor in bacteria. The sequence corresponding to 5.8S rRNA forms the 5′ end of the large (23S) rRNA, that is, there is no processing between these sequences. **Figure 24.38** shows that the precursor also contains the 5S rRNA and one or two tRNAs. In *E. coli*, the 7 *rrn* operons are

dispersed around the genome; four *rrn* loci contain one tRNA gene between the 16S and 23S rRNA sequences, and the other *rrn* loci contain two tRNA genes in this region. Additional tRNA genes may or may not be present between the 5S sequence and the 3′ end. So the processing reactions required to release the products depend on the content of the particular *rrn* locus.

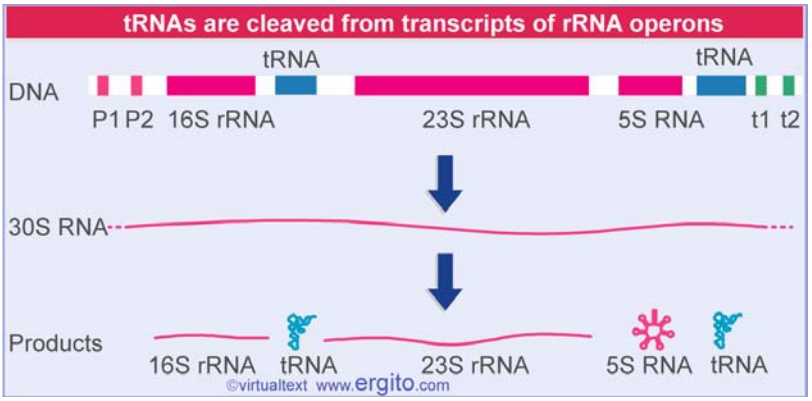

Figure 24.38 The *rrn* operons in *E. coli* contain genes for both rRNA and tRNA. The exact lengths of the transcripts depend on which promoters (P) and terminators (t) are used. Each RNA product must be released from the transcript by cuts on either side.

In both prokaryotic and eukaryotic rRNA processing, ribosomal proteins (and possibly also other proteins) bind to the precursor, so that the substrate for processing is not the free RNA but is a ribonucleoprotein complex.

Reviews

980. Venema, J. and Tollervey, D. (1999). *Ribosome synthesis in S. cerevisiae.* Annu. Rev. Genet. 33, 261-311.

RNA SPLICING AND PROCESSING

5.24.22 Small RNAs are required for rRNA processing

Key Terms

A **snoRNA** is a small nuclear RNA that is localized in the nucleolus.

Key Concepts

- The C/D group of snoRNAs is required for modifying the 2′ position of ribose with a methyl group.
- The H/ACA group of snoRNAs is required for converting uridine to pseudouridine.
- In each case the snoRNA base pairs with a sequence of rRNA that contains the target base to generate a typical structure that is the substrate for modification.

Processing and modification of rRNA requires a class of small RNAs called **snoRNAs** (small nucleolar RNAs). There are 71 snoRNAs in the yeast (*S. cerevisiae*) genome. They are associated with the protein fibrillarin, which is an abundant component of the nucleolus (the region of the nucleus where the rRNA genes are transcribed). Some snoRNAs are required for cleavage of the precursor to rRNA; one example is U3 snoRNA, which is required for the first cleavage event in both yeast and *Xenopus* (740). We do not know what role the snoRNA plays in cleavage. It could be required to pair with the rRNA sequence to form a secondary structure that is recognized by an endonuclease.

Two groups of snoRNAs are required for the modifications that are made to bases in the rRNA (see *Great Experiments 4.6 Small nucleolar RNAs guide rRNA modification*). The members of each group are identified by very short conserved sequences and common features of secondary structure (1216; 1217).

The C/D group of snoRNAs is required for adding a methyl group to the 2′ position of ribose. There are >100 2′-O-methyl groups at conserved locations in vertebrate rRNAs. This group takes its name from two short conserved sequences motifs called boxes C and D. Each snoRNA contains a sequence near the D box that is complementary to a region of the 18S or 28S rRNA that is methylated. Loss of a particular snoRNA prevents methylation in the rRNA region to which it is complementary.

Figure 24.39 suggests that the snoRNA base pairs with the rRNA to create the duplex region that is recognized as a substrate for methylation. Methylation occurs within the region of complementarity, at a position that is fixed 5 bases on the 5′ side of the D box (741; 1220). Probably each methylation event is specified by a different snoRNA; ~40 snoRNAs have been characterized so far. The methylase(s) have not been characterized; one possibility is that the snoRNA itself provides part of the methylase activity.

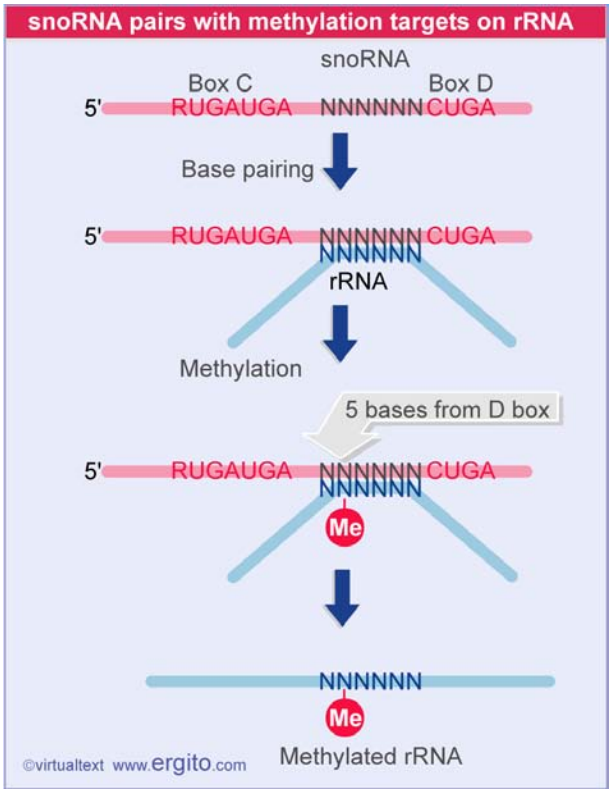

Figure 24.39 A snoRNA base pairs with a region of rRNA that is to be methylated.

Another group of snoRNAs is involved in the synthesis of pseudouridine. There are 43 ψ residues in yeast rRNAs and ~100 in vertebrate rRNAs. The synthesis of pseudouridine involves the reaction shown in **Figure 24.40** in which the N1 bond from uridylic acid to ribose is broken, the base is rotated, and C5 is rejoined to the sugar.

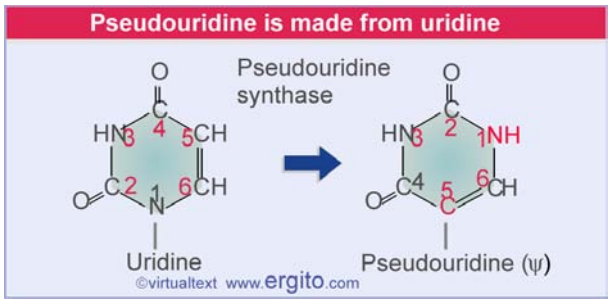

Figure 24.40 Uridine is converted to pseudouridine by replacing the N1-sugar bond with a C5-sugar bond and rotating the base relative to the sugar.

Pseudouridine formation in rRNA requires the H/ACA group of ~20 snoRNAs. They are named for the presence of an ACA triplet 3 nucleotides from the 3′ end and a partially conserved sequence (the H box) that lies between two stem-loop hairpin structures. Each of these snoRNAs has a sequence complementary to rRNA within the stem of each hairpin. **Figure 24.41**shows the structure that would be produced by

pairing with the rRNA. Within each pairing region, there are two unpaired bases, one of which is a uridine that is converted to pseudouridine (742; 1218).

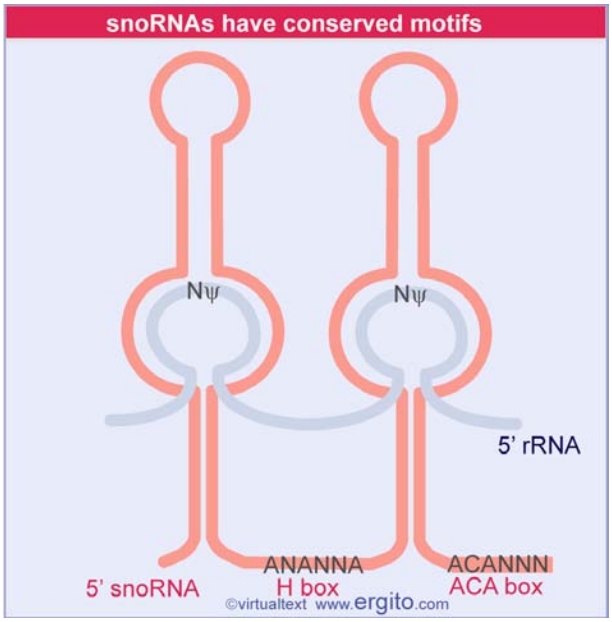

Figure 24.41 H/ACA snoRNAs have two short conserved sequences and two hairpin structures, each of which has regions in the stem that are complementary to rRNA. Pseudouridine is formed by converting an unpaired uridine within the complementary region of the rRNA.

The H/ACA snoRNAs are associated with a nucleolar protein called Gar1p, which is required for pseudouridine formation, but its function is unknown (1219). The known pseudouridine synthases are proteins that function without an RNA cofactor. Synthases that could be involved in snoRNA-mediated pseudouridine synthesis have not been identified.

The involvement of the U7 snRNA in 3′ end generation, and the role of snoRNAs in rRNA processing and modification, is consistent with the view we develop in *Molecular Biology 5.26 Catalytic RNA* that many – perhaps all – RNA processing events depend on RNA-RNA interactions. As with splicing reactions, the snRNA probably functions in the form of a ribonucleoprotein particle containing proteins as well as the RNA. It is common (although not the only mechanism of action) for the RNA of the particle to base pair with a short sequence in the substrate RNA.

Last updated on 10-30-2000

References

740. Kass, S. et al. (1990). *The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing*. Cell 60, 897-908.

741. Kiss-Laszlo, Z. et al. (1996). *Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs*. Cell 85, 1077-1068.

742. Ni, J., Tien, A. L., and Fournier, M. J. (1997). *Small nucleolar RNAs direct site-specific synthesis of pseudouridine in rRNA*. Cell 89, 565-573.

1216. Balakin, A. G., Smith, L., and Fournier, M. J. (1996). *The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions.* Cell 86, 823-834.

1217. Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997). *The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation.* Genes Dev. 11, 941-956.

1218. Ganot, P., Bortolin, M. L., and Kiss, T. (1997). *Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs*. Cell 89, 799-809.

1219. Bousquet-Antonelli, C., Henry, Y., G'elugne, J. P., Caizergues-Ferrer, M., and Kiss, T. (1997). *A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs.* EMBO J. 16, 4770-4776.

1220. Kiss-Laszlo, Z., Henry, Y., and Kiss, T. (1998). *Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA.* EMBO J. 17, 797-807.

RNA SPLICING AND PROCESSING

5.24.23 Summary

Splicing accomplishes the removal of introns and the joining of exons into the mature sequence of RNA. There are at least four types of reaction, as distinguished by their requirements *in vitro* and the intermediates that they generate. The systems include eukaryotic nuclear introns, group I and group II introns, and tRNA introns. Each reaction involves a change of organization within an individual RNA molecule, and is therefore a *cis*-acting event.

pre-mRNA splicing follows preferred but not obligatory pathways. Only very short consensus sequences are necessary; the rest of the intron appears irrelevant. All 5′ splice sites are probably equivalent, as are all 3′ splice sites. The required sequences are given by the GU-AG rule, which describes the ends of the intron. The UACUAAC branch site of yeast, or a less well conserved consensus in mammalian introns, is also required. The reaction with the 5′ splice site involves formation of a lariat that joins the GU end of the intron via a 5′–2′ linkage to the A at position 6 of the branch site. Then the 3′–OH end of the exon attacks the 3′ splice site, so that the exons are ligated and the intron is released as a lariat. Both reactions are transesterifications in which bonds are conserved. Several stages of the reaction require hydrolysis of ATP, probably to drive conformational changes in the RNA and/or protein components. Lariat formation is responsible for choice of the 3′ splice site. Alternative splicing patterns are caused by protein factors that either stimulate use of a new site or that block use of the default site.

pre-mRNA splicing requires formation of a spliceosome, a large particle that assembles the consensus sequences into a reactive conformation. The spliceosome most often forms by the process of intron definition, involving recognition of the 5′ splice site, branch site, and 3′ splice site. An alternative pathway involves exon definition, which involves initial recognition of the 5′ splice sites of both the substrate intron and the next intron. Its formation passes through a series of stages from the E (commitment) complex that contains U1 snRNP and splicing factors, through the A and B complexes as additional components are added.

The spliceosome contains the U1, U2, U4/U6, and U5 snRNPs and some additional splicing factors. The U1, U2, and U5 snRNPs each contain a single snRNA and several proteins; the U4/U6 snRNP contains 2 snRNAs and several proteins. Some proteins are common to all snRNP particles. The snRNPs recognize consensus sequences. U1 snRNA base pairs with the 5′ splice site, U2 snRNA base pairs with the branch sequence, U5 snRNP acts at the 5′ splice site. When U4 releases U6, the U6 snRNA base pairs with U2, and this may create the catalytic center for splicing. An alternative set of snRNPs provides analogous functions for splicing the U12-dependent subclass of introns. The snRNA molecules may have catalytic-like roles in splicing and other processing reactions.

In the nucleolus, two groups of snoRNAs are responsible for pairing with rRNAs at sites that are modified; group C/D snoRNAs indicate target sites for methylation, and group ACA snoRNAs identify sites where uridine is converted to pseudouridine.

Splicing is usually intramolecular, but *trans*-(intermolecular) splicing occurs in trypanosomes and nematodes. It involves a reaction between a small SL RNA and the pre-mRNA. The SL RNA resembles U1 snRNA and may combine the role of providing the exon and the functions of U1. In worms there are two types of SL RNA, one used for splicing to the 5′ end of an mRNA, the other for splicing to an internal site.

Group II introns share with nuclear introns the use of a lariat as intermediate, but are able to perform the reaction as a self-catalyzed property of the RNA. These introns follow the GT-AG rule, but form a characteristic secondary structure that holds the reacting splice sites in the appropriate apposition.

Yeast tRNA splicing involves separate endonuclease and ligase reactions. The endonuclease recognizes the secondary (or tertiary) structure of the precursor and cleaves both ends of the intron. The two half-tRNAs released by loss of the intron can be ligated in the presence of ATP.

The termination capacity of RNA polymerase II has not been characterized, and 3′ ends of its transcripts are generated by cleavage. The sequence AAUAAA, located 11-30 bases upstream of the cleavage site, provides the signal for both cleavage and polyadenylation. An endonuclease and the poly(A) polymerase are associated in a complex with other factors that confer specificity for the AAUAAA signal.