10 - RNA Modifications

After the RNA molecule is produced by transcription (Part 9), the structure of the RNA is often modified prior to being translated into a protein product. These **RNA modifications** apply mainly to eukaryotic RNA transcripts.

Key Questions

- Which group of organisms modify their RNA transcripts?

Overview of RNA Modifications

The modifications to eukaryotic RNA transcripts include the following:

- **5’ end capping.** 5’ end capping involves the attachment of a modified nucleotide called **7-methylguanosine (7-mG)** to the 5’ end of RNA molecules. The added 7-mG is sometimes called the **5’ cap**.
- **3’ end polyadenylation.** 3’ end polyadenylation involves the addition of a string of adenine (A) nucleotides to the 3’ end of the RNA molecule. The added sequence of A nucleotides is called the **polyA tail**.
- **RNA splicing.** Most eukaryotic genes are **split genes,** being composed of both intron DNA sequences and exon sequences. For split genes, initial transcription in the nucleus produces a **precursor mRNA (pre-mRNA)** molecule. This pre-mRNA then goes through 5' end capping, 3' end polyadenylation, and finally RNA splicing. During RNA splicing, the **intron** sequences are removed from the pre-mRNA and discarded (see **Figure 10.1**). The remaining exon RNA segments are spliced together to produce a **mature mRNA** molecule that is transported to the cytoplasm of the cell for translation.
- **RNA processing.** RNA processing involves cutting larger RNA transcripts into smaller ones. RNA processing involves both **exonucleases** (removing nucleotides from the ends of the RNA transcript) or **endonucleases** (cleaving the RNA transcript at an internal site). The ribosomal RNA (rRNA) molecules that are essential components within ribosomes (see Part 11) experience RNA processing after transcription.
- **RNA editing.** RNA editing involves changing the nucleotide sequence of the mRNA molecule prior to translation.
- **Base modification.** During base modification, nitrogenous bases within the RNA transcript are covalently modified by the addition of chemical groups, such as methyl groups.

The remainder of Part 10 will focus on the first three RNA modifications: 5' end capping, 3' end polyadenylation, and RNA splicing.

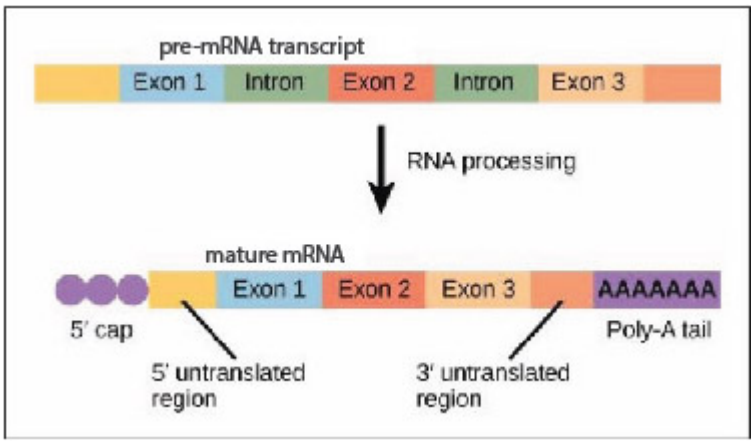

Figure 10.1 ***RNA Modifications Overview*** *--- Image used from OpenStax (access for free at https://openstax.org/books/biology-2e/pages/1-introduction*

Key Questions

- What is the difference between a pre-mRNA and a mature mRNA molecule?
- What is the difference between an intron and an exon DNA sequence?
- What is meant by 5′ end capping?
- What is meant by 3′ end polyadenylation?

5′ End Capping

The 5′ end of the pre-mRNA molecule is modified by the addition of a **7-methylguanosine (7-mG)** nucleotide. The process of adding the 7-mG to the pre-mRNA is **5′ end capping**. 5′ end capping is the first RNA modification, occurring as soon as the 5' end of the pre-mRNA emerges from RNA polymerase II during transcription. 5′ end capping (see **Figure 10.2**) involves the following enzymes:

1. **RNA 5′-triphosphatase.** Recall that RNA polymerases do not require a primer to initiate transcription (see Part 9). As a result, the first nucleotide incorporated into the RNA has three phosphate groups attached to the 5' carbon. RNA 5'-triphosphatase removes one of the three phosphates from this nucleotide.
2. **Guanylyltransferase.** Guanylyltransferase cleaves GTP to produce GMP and pyrophosphate (PP_i). Guanylyltransferase then attaches the phosphate group of the GMP molecule to the two phosphate groups on the nucleotide at the 5′ end of the pre-mRNA transcript. It is important to note that an unusual 5′ to 5′ linkage is formed, placing three phosphate groups between the 5' carbons on two adjacent nucleotides.
3. **Methyltransferase.** Methyltransferase attaches a methyl group to the added guanine nitrogenous base, resulting in the 7-mG cap.

The 7-mG cap on eukaryotic mRNAs has at least three functions. The 7-mG cap:

- Serves as a binding site for proteins that transport the mRNA from the nucleus to the cytoplasm of the cell.
- Serves as a recognition site for translation factor proteins that help the ribosome bind to the mRNA. Once the ribosome binds to the mRNA, translation begins (see Part 11).
- Protects the 5′ end of the mRNA transcript from exonuclease digestion.

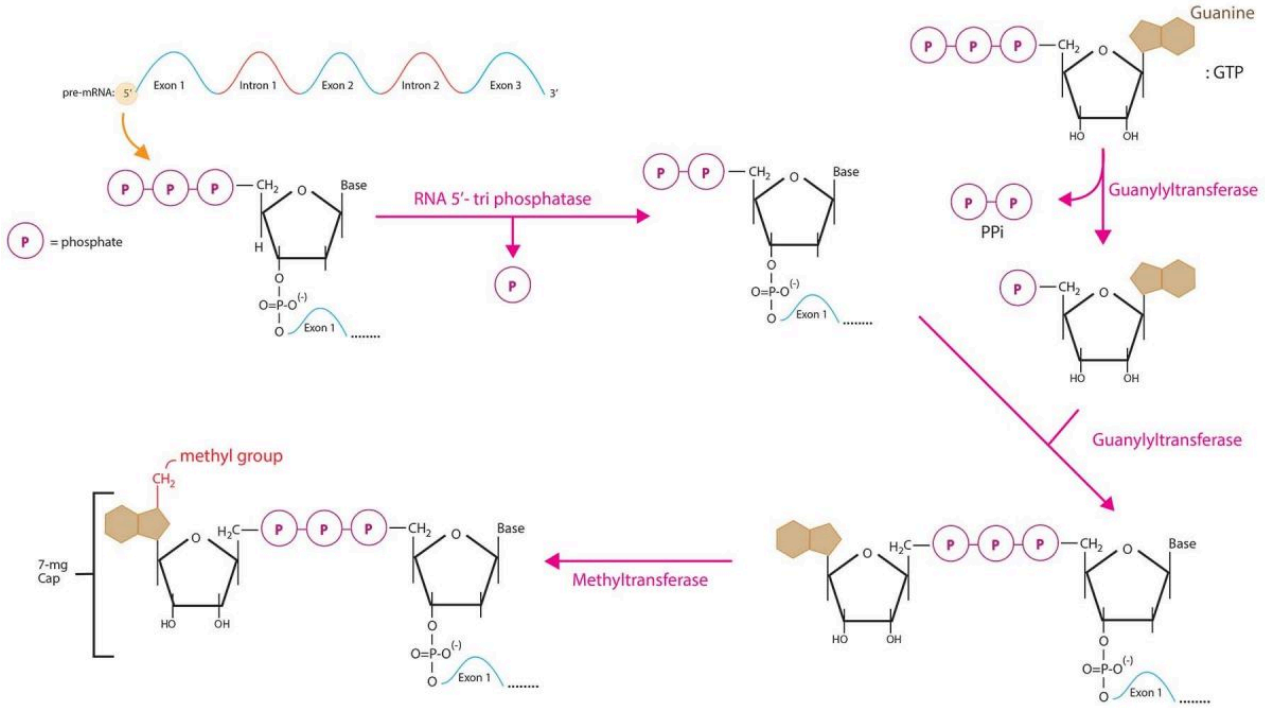

Figure 10.2 ***5' end capping mechanism*** *--- Image created by JET*

Key Questions

- How does the 7-mG structure contribute to translation?
- Which nucleotide triphosphate provides the energy for 5' end capping?
- What is unusual about the covalent bonds between 7-mG and the rest of the pre-mRNA molecule?

3' End Polyadenylation

The 3' end of the pre-mRNA is modified by the addition of a **polyA tail**, a string of approximately 250 adenine (A) nucleotides. The process of adding a polyA tail to the mRNA transcript (see **Figure 10.3**), called **3' end polyadenylation**, involves:

1. The detection of two recognition sequences (**polyadenylation signal sequences**) near the 3' end of the pre-mRNA molecule. The first polyadenylation signal sequence, 5'-AAUAAA-3', is recognized by the endonuclease **cleavage and polyadenylation specificity factor (CPSF)** protein. The second polyadenylation signal sequence, enriched in guanine and uracil bases, is called the GU-rich sequence. This GU-rich sequence is the binding site for the **cleavage stimulatory factor (CstF)** protein. When CstF and CPSF bind to their respective polyadenylation signal sequences, CstF activates CPSF.
2. The CPSF protein cleaves the pre-mRNA between the two polyadenylation signal sequences. When CPSF cleaes the pre-mRNA molecule, the pre-mRNA is released from RNA polymerase II. The new 3' end of the pre-mRNA is then available for the addition of a polyA tail.
3. **Poly(A)-polymerase (PAP)** attaches 250 adenine nucleotides to the newly generated 3' end of the pre-mRNA transcript. PAP is an unusual RNA polymerase that does not require a template and only forms phosphodiester bonds between adenine nucleotides.

The polyA tail on the mRNA has at least three functions. The polyA tail functions to:

- Protect the 3’ end of the pre-mRNA transcript from exonuclease degradation.
- Promote the transport of the mRNA from the nucleus to the cytoplasm of the cell.
- Help the ribosome bind to the mRNA to initiate translation.

The 3’ end polyadenylation process occurs after 5’ end capping, but prior to RNA splicing. In fact, 3’ end polyadenylation assists in terminating transcription in eukaryotes by the torpedo model (see Part 9).

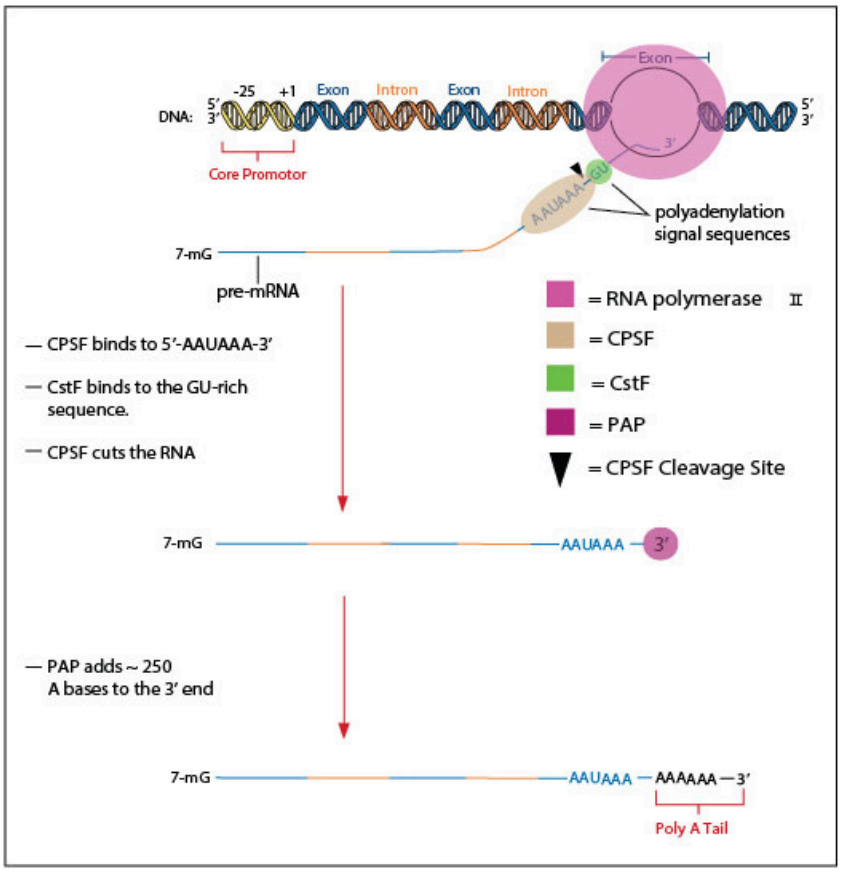

Figure 10.3 ***3’ end polyadenylation mechanism*** *--- Image created by SL*

Key Questions

- How does 3’ end polyadenylation contribute to transcription termination in eukaryotes?
- Describe the functions of CPSF, CstF, and Poly(A)-polymerase during 3' end polyadenylation.

Splicing of Group I and Group II Introns

There are three general mechanisms used by eukaryotes to remove introns from RNA molecules. The **group I** and **group II** mechanisms are limited to certain types of eukaryotes or certain organelles within a eukaryotic cell. For example, the group I mechanism removes the introns found in ribosomal RNA (rRNA) molecules in certain protozoa. The group II mechanism removes the introns found in the mRNA and transfer RNA (tRNA) molecules produced by mitochondrial and chloroplast genes. The **spliceosome mechanism** is the major mechanism that is used to remove introns from pre-mRNA transcripts in the nucleus of eukaryotic cells.

- **Removing group I introns.** RNA splicing of group I introns occurs by **self-splicing**, meaning that the precursor RNA molecule catalyzes the removal of its own intron (see **Figure 10.4**). These catalytic precursor RNA molecules are examples of a unique group of molecules classified as RNA enzymes **(ribozymes)**. In fact, the discovery of the Group I ribozyme was the first demonstration that a molecule other than a protein could serve as an enzyme. The ribozyme mechanism to remove group I introns occurs as follows:

    1. A free **guanosine nucleoside** (guanine nitrogenous base covalently linked to a ribose sugar; no phosphate groups) binds to a pocket within the intron. The guanosine nucleoside bound to the intron serves as an enzyme cofactor (i.e., assists the ribozyme in catalysis) for the remaining steps in the reaction.
    2. A break forms at the junction between the 3' end of the first exon and the 5' end of the intron.
    3. The released 3' end of the first exon then cleaves the phosphodiester bond between the 3' end of the intron and the 5' end of the second exon.
    4. A phosphodiester bond is formed that links the first and second exons together, generating a mature RNA molecule. The intron is released and degraded.

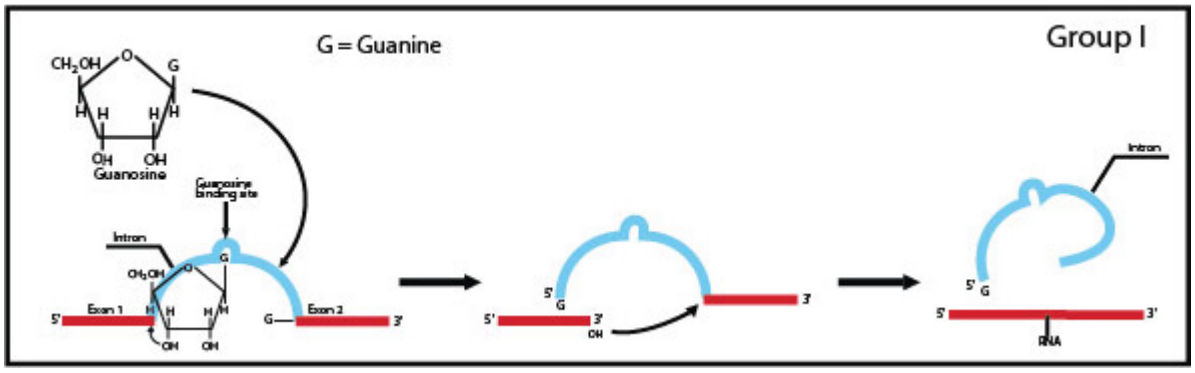

Figure 10.4 ***Removing group I introns*** *--- Image created by SL*

- **Removing group II introns.** RNA splicing of group II introns also occurs by self-splicing, meaning that the precursor RNA is an RNA enzyme (ribozyme) that removes its own intron (see **Figure 10.5**). The self-splicing of group II introns involves:

    1. The 2'-OH group of an **adenine nucleotide** within the intron cleaves the phosphodiester bond between the 3' end of the first exon and the 5' end of the intron. In this reaction, the adenine nucleotide serves as an enzyme cofactor for the reaction.
    2. The released 3' end of the first exon then cleaves the phosphodiester bond between the 3' end of the intron and the 5' end of the second exon.
    3. A phosphodiester bond is formed that links the first and second exons, generating a mature RNA transcript. The intron is released and degraded.

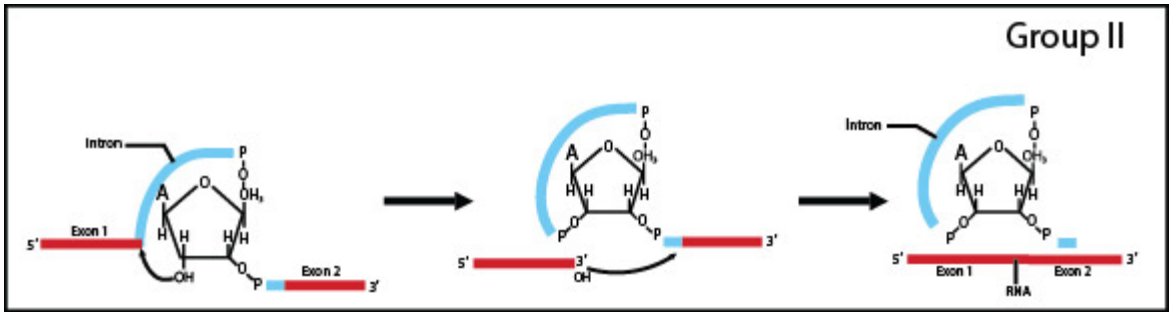

Figure 10.5 ***Removing group II introns*** *--- Image created by SL*

Key Questions

- What is a ribozyme?
- Describe the major events that occur in the Group I and Group II splicing mechanisms.
- What molecules serve as enzyme cofactors in the Group I and Group II splicing mechanisms?

Removal of Introns by Spliceosomes

Transcription of most structural genes in the nucleus of eukaryotic cells produces pre-mRNA molecules; the removal of the introns within these pre-mRNA molecules involves a large multi-subunit **spliceosome** complex. To remove introns from the pre-mRNA, the spliceosome binds to recognition sequences within the intron (see **Figure 10.6**). These **intron recognition sequences** include:

- **The 5′ splice site.** The 5′ splice site is a 5′-GU-3′ RNA sequence at the 5′ end of the intron.
- **The branch site.** The branch site is an adenine nucleotide (A) near the middle of the intron RNA sequence.
- **The 3′ splice site.** The 3′ splice site is an 5′-AG-3′ RNA sequence at the 3′ end of the intron.

The spliceosome complex contains multiple subunits; these subunits are called **small nuclear ribonucleoproteins** or **snRNPs** (“snurps”). Each snRNP within the spliceosome complex is composed of a **small nuclear RNA (snRNA)** molecule that acts as an RNA enzyme (ribozyme) to remove the introns from the pre-mRNA molecule. snRNPs are also composed of proteins that function to stabilize snRNP structure.

The spliceosome splicing mechanism occurs as follows:

1. The **U1** snRNP binds to the 5′ splice site within the intron RNA sequence, while the **U2** snRNP binds to the branch site adenine within the intron.
2. Additional snRNPs called **U4, U5**, and **U6** bind to the intron. These five snRNPs (U1, U2, U4, U5, and U6) form the spliceosome complex.
3. The intron forms a loop bringing the two exon sequences to be linked close together.
4. The 5′ splice site within the intron is cut by U1, and the 5′ end of the intron is covalently linked to the 2′-OH group of the branch site adenine, forming an RNA loop structure called a **lariat**.
5. The U1 and U4 snRNPs are released.
6. The 3′ splice site within the intron is cut by the U5 snRNP.
7. A phosphodiester bond is formed that links the two exons together to form the mature mRNA molecule.
8. The intron is released along with the U2, U5, and U6 snRNPs.

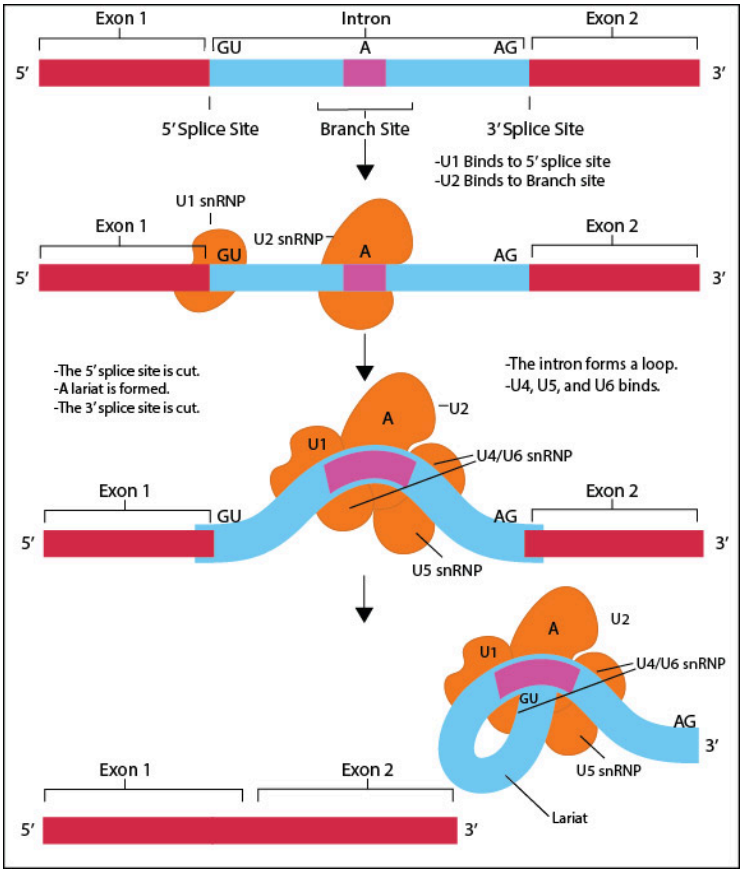

Figure 10.6 ***Spliceosome splicing*** *--- Image created by SL*

Key Questions

- Which two splicing mechanisms are found in human cells?
- What are the names of the three RNA sequences found within introns removed by the spliceosome?
- Which snRNP component is the ribozyme?
- What are the functions of the U1 and U5 snRNPs?

Identifying Introns Using R-Loop Experiments

Introns were initially identified within eukaryotic genes by performing **R-loop (hybridization) experiments**. These R-loop experiments relied on separating the two DNA strands within a gene, allowing a mRNA molecule to form hydrogen bonds (hybridize) with the template DNA strand, and adding the coding strand DNA, which attempts to form hydrogen bonds with the template DNA strand. Finally, the resulting nucleic acid structure was examined in an electron microscope. Below are the results expected from two R-loop experiments, one experiment involving the pre-mRNA (before RNA modifications), the other experiment involving the mature mRNA (after RNA modifications).

- **Gene hybridized to the pre-mRNA.** The pre-mRNA forms hydrogen bonds with the template DNA strand preventing the coding DNA strand from binding. Because the coding DNA strand fails to bind to the template DNA strand, the coding DNA strand extends outward from the RNA-DNA hybrid region. This loop where the coding DNA strand cannot bind to the template DNA strand is called an RNA displacement loop or **R loop** (see **Figure 10.7A**).
- **Gene hybridized to the mature mRNA.** Hybridization between the template DNA strand and the mature mRNA forces the intron DNA sequences in the template DNA strand to loop out, because the mature mRNA lacks intron sequences. Adding the coding DNA strand produces R-loops with intervening regions of double-stranded DNA (i.e., the intron sequences within the template and the coding DNA strands form hydrogen bonds) called **intron loops** (see **Figure 10.7B**).

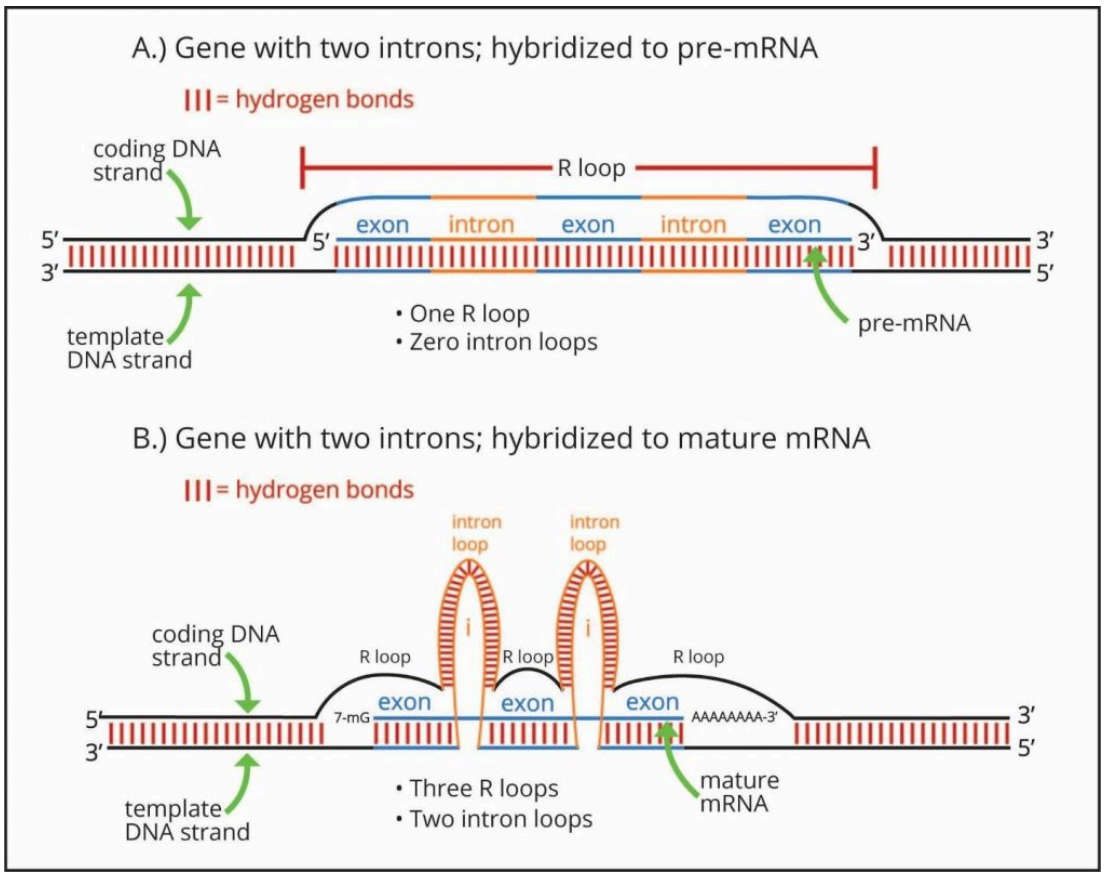

Figure 10.7 ***R-Loop Results*** *--- Image created by SL*

Key Questions

- Suppose a gene contains four introns and is hybridized with its mature mRNA. How many R loops would be observed in the electron microscope at the end of an R-loop experiment? How many intron loops would be observed?

Identifying Introns by Comparing gDNA with cDNA

Introns within genes can also be identified by comparing the length of a **genomic DNA (gDNA)** version of a gene to the **complementary DNA (cDNA)** version of the same gene. gDNA is the version of a gene found in the genome; the gDNA version of a gene contains both introns and exons. cDNA is produced in the laboratory by **reverse transcription** (see Part 8). Reverse transcription converts mature mRNA into a cDNA molecule using the viral enzyme **reverse**

transcriptase. Since the cDNA molecule is produced from the mature mRNA, cDNA molecules contain exons but lack introns. The gDNA version of the gene, which contains introns, will be longer than the cDNA version of the same gene, which lacks introns.

The **polymerase chain reaction (PCR)** technique (see Part 8) can be used to make billions of copies of the gDNA and the cDNA versions of any gene of interest. The gDNA and cDNA PCR products are then separated by size using **agarose gel electrophoresis** (see Part 8). The size difference between the gDNA and the cDNA copy of the gene can be easily observed on an agarose gel (see **Figure 3.2**).

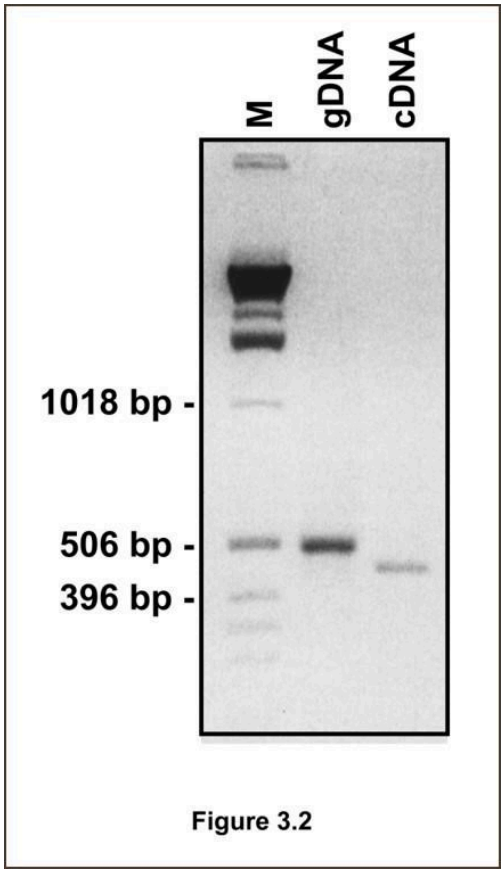

Figure 10.8 ***Comparing gDNA to cDNA to identify* introns.** *The gene in question contains a 75 base pair (bp) intron. --- Image provided by K. Mark DeWall*

Key Questions

- What is the difference between gDNA and cDNA?
- How can comparing gDNA to cDNA on an agarose gel help you indentify an intron?

Alternative Splicing

Alternative splicing involves splicing a single type of pre-mRNA molecule in different ways to produce multiple mature mRNA molecules (see **Figure 10.9**). Each of these mature mRNAs can then produce slightly different proteins upon translation. These distinct, yet related protein **isoforms**, all derived from a single gene, can have specialized functions. Alternative splicing is beneficial in that it allows eukaryotes to carry fewer genes in the genome, permitting a relatively small number of genes the flexibility to encode a vast array of proteins. In humans, it is estimated that 30–60% of the

genes in the genome are alternatively spliced. As a result, the human genome, which contains approximately 23,000 structural genes, can produce at least ten times that number of unique protein products.

One example of alternative splicing involves the human α-tropomyosin gene, a gene involved in muscle contraction. The α-tropomyosin gene contains 14 exons and 13 introns. The α-tropomyosin gene contains two types of exons:

- **Constitutive exons.** Constitutive exons are included in all of the α-tropomyosin mature mRNAs products of alternative splicing. These exons likely encode amino acid sequences that maintain the general three-dimensional structure of the encoded α-tropomyosin protein.
- **Alternative exons.** Alternative exons vary between α-tropomyosin mature mRNAs. In one cell type, one combinations of alternative exons are spliced together with the constitutive exons to make a mature mRNA molecule. In another cell type, a different combination of alternative exons are spliced together with the constitutive exons to make a mature mRNA molecule. The result is two related proteins that have slightly different functions to meet the unique needs of these two different cell types.

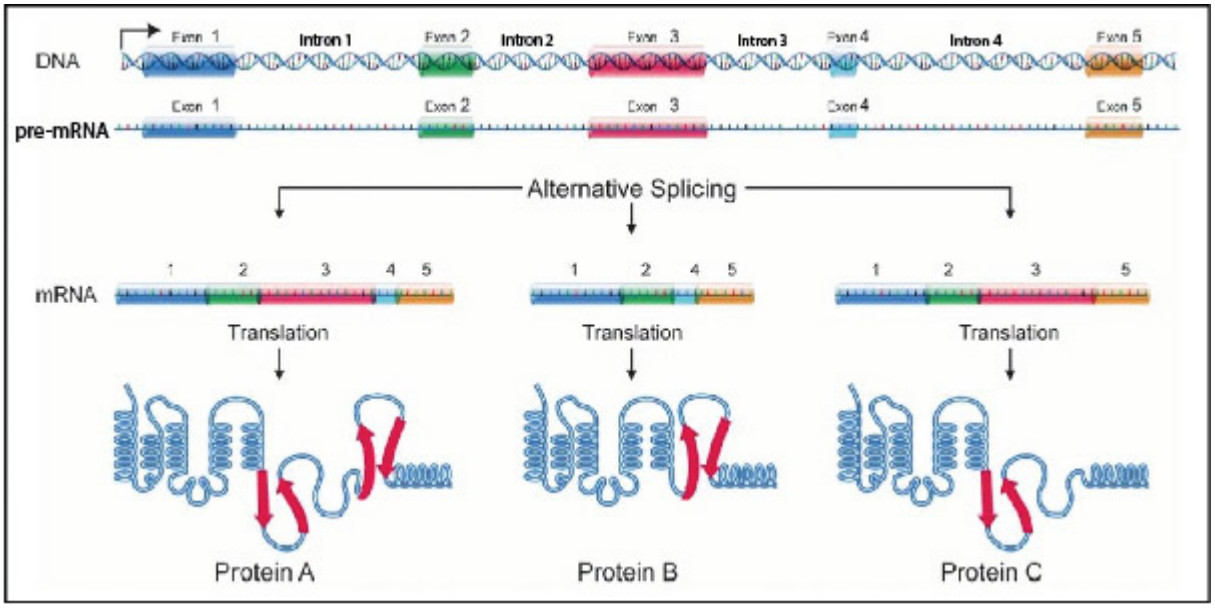

Figure 10.9 ***Alternative splicing allows one gene to produce three proteins.*** *In this example, exons 1, 2, and 5 are constitutive exons, while exons 3 and 4 are alternative exons. --- DNA Alternative Splicing by National Human Genome Research Institue and is used under CC0*

Key Questions

- Why is alternative splicing advantageous?
- What are protein isoforms?
- What is the difference between a constitutive exon and an alternative exon?

Patterns of Alternative Splicing

Alternative splicing is regulated by **splicing factor** proteins. These splicing factor proteins help the spliceosome choose which intron splice sites to cut during RNA splicing. Each cell type has a different collection of splicing factor proteins, allowing different RNA splicing patterns to occur in each cell type.

Here are some common alternative splicing patterns observed in eukaryotic cells:

- **Exon Skipping.** Some splicing factor proteins act as **splice repressors**. Splice repressor proteins prevent the spliceosome from recognizing a particular 3' splice site within an intron (see Figure 10.10). When a splice repressor protein blocks a 3′ splice site within an intron, the 3′ splice site in the next intron is chosen for splicing instead, and the intervening exon is removed from the pre-mRNA molecule (**exon skipping**).
- **Alternative 5' and 3' Splice Sites.** In addition to the 5′ splice site, the branch site, and the 3′ splice sites discussed earlier, there are other pre-mRNA sequences involved in RNA splicing. These additional sequence elements, often located within a nearby exon, can promote the use of a particular 5′ or 3′ splice site. For example, some potential 5′ or 3′ splice sites in the pre-mRNA are poorly recognized by the spliceosome. In certain cell types, the binding of a **splice activator protein** to a **splice enhancer sequence** within a nearby exon promotes the use of these otherwise poorly recognized 5′ or 3′ splice sites (see **Figure 10.10**). When a splice activator protein binds to a splice enhancer sequence, an exon is included in the mature mRNA (i.e., the exon is not skipped).
- **Mutually Exclusive Exons.** In some cases, splicing events are coordinated between different cell types to ensure that unique protein isoforms are produced by each cell type. For example, suppose there are four exons (three introns) in a pre-mRNA molecule. During splicing in one cell type, exon two is consistently retained in the mature mRNA, while exon three is spliced out. In a different cell type, exon two is always spliced out, while exon three is.01233 retained in the mature mRNA. Exons one and four are found in the mature mRNAs in both cell types and are thus constitutive exons.

Scientists are still learning the true complexity of alternative splicing. It appears that alternative splicing patterns are cell-type and developmental stage specific. Moreover, mutations often lead to aberrant splicing patterns. This aberrant splicing produces abnormal protein isoforms and in some cases, disease phenotypes.

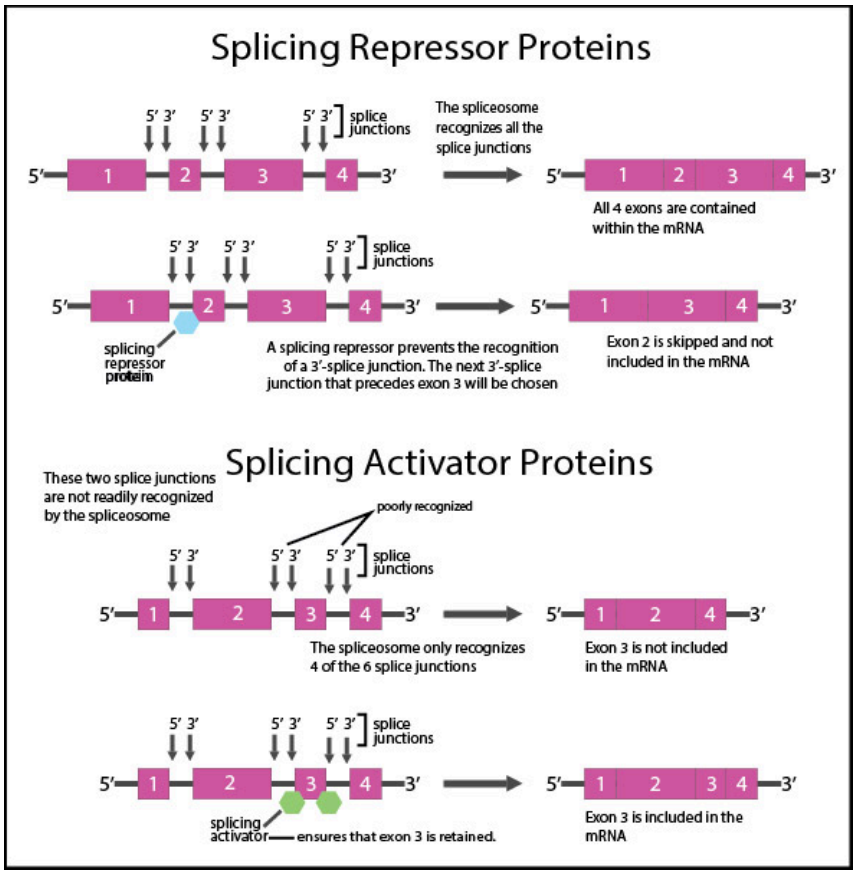

Figure 10.10 ***Splicing repressor and activator proteins*** *--- Image created by SL*

Key Questions

- What happens when a splice repressor protein binds to the 3’ splice site within an intron?
- What effect would a splice activator protein binding to a splice enhancer sequence have on alternative splicing?
- What is meant by the term mutually exclusive exons?

Review Questions

Fill in the blank:

1. _____ is an endonuclease that releases the pre-mRNA from RNA polymerase II to terminate transcription.
2. _____ is an enzyme that attaches two nucleotides together via a 5’ to 5’ linkage.
3. One function of the 7-mG cap is to _____.
4. A _____ protein prevents the spliceosome from binding to a 3’ splice site.
5. _____ is an enzyme that adds adenine nucleotides to the 3' end of a pre-mRNA. These adenine nucleotides are added in the 5' to 3' direction.
6. The Group I intron splicing mechanism uses the nucleoside _____ as a cofactor during catalysis, while the _____ intron splicing mechanism uses an adenine nucleotide as a cofactor during catalysis.
7. The U2 snRNP binds to the _____ site of the pre-mRNA.
8. Spliceosome subunits are composed of two components: proteins and _____.
9. _____ is a pattern of alternative splicing in which an exon is always retained in one cell while that same exon is always skipped in another cell.