Chapter 3: Transcription and mRNA Modification

Have you ever wondered why the tissue of your eye is different from the tissue of the liver or kidney? Have you ever looked at your skin and thought why does it look the way it does or function the way it does? How is it that every cell that has the same DNA, but can function so differently from another cell? We can begin to answer these questions by learning about a central process in cells called transcription.

# Transcription Starts the Process of Gene Expression

Consider that all of the cells in a multicellular organism have arisen by division from a single fertilized egg and therefore, all have the same DNA. Division of that original fertilized egg produces, in the case of humans, over a trillion cells by the time a baby is produced from that egg (that's a lot of DNA replication!). Yet, we also know that a baby is not a giant ball of a trillion identical cells, but has many different kinds of cells that make up different tissues such as skin, muscle, bone, and nerves. How can cells that have identical DNA look so different and have different functions?

The answer lies in gene expression regulation, which is the process of using the information stored in DNA to generate different proteins through the steps of the central dogma of biology. Although all of the cells in a baby have the same DNA, each different cell type uses a unique subset of the genes in the DNA to direct the synthesis of a distinctive set of RNAs and proteins. The first step in gene expression is transcription.

What is transcription? **Transcription** is the process of copying information from DNA sequences into RNA sequences. This process is also known as DNA-dependent RNA synthesis. When a sequence of DNA is transcribed, only one portion of one of the two DNA strands is copied into RNA; this is contrasted with the process of DNA replication (Chapter 14) where both strands of DNA must be copied, and in their entirety

Transcription copies short stretches of the coding regions (DNA sequence that is eventually made into proteins) of DNA to generate RNA. In humans, these short stretches can average about 8.5 kilobases (kb) and consist of both exons and introns. Exons contain the information to make proteins and introns are non-coding, intervening, DNA sequences. Different genes may be copied into RNA at varying times in the cell's life-cycle. RNAs are temporary copies of instructions of the information in DNA and different sets of instructions are copied for use at different times.

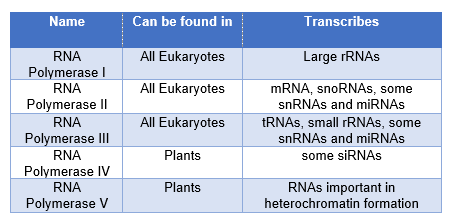
**Table 3-1: General features of transcription.**

|  |
| --- |
| General Features of Transcription |
| * The process of making RNA copies using a DNA as template is known as transcription * The process of transcription produces all sorts of RNAs (mRNA, tRNA, rRNA, etc) * One strand of the DNA serves as a template for the synthesis of RNA * Enzymes that synthesize RNA are called RNA polymerases * RNA polymerases synthesize RNA in the 5’ to 3’ direction * RNA polymerases do not need a primer * RNA polymerases uses rNTPs (ATP, GTP, UTP, and CTP) to build the new RNA strand * RNA polymerases bind at specific DNA sequences called promoters to start transcription * RNA polymerases stop RNA synthesis when they reach sequences called terminators |

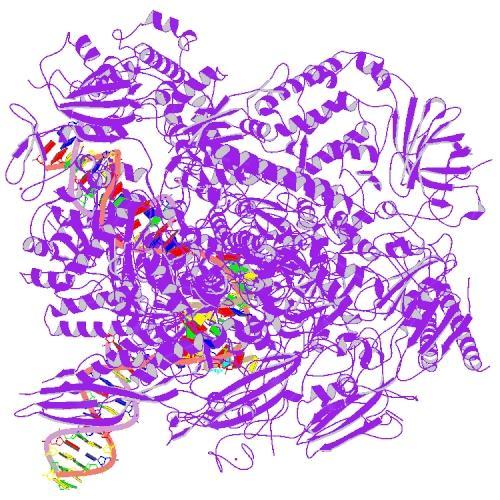
Cells make several different kinds of RNA:

* **Messenger RNA**: mRNAs that code for proteins
* **Ribosomal RNA**: rRNAs that form part of ribosomes
* **Transfer RNA**: tRNAs that carry the amino acids to the ribosome during translation
* **MicroRNA**s: [miRNAs](https://en.wikipedia.org/wiki/MicroRNA) that regulate gene expression
* Many other small RNAs that have a variety of unique functions, often, but not always, playing a role in regulating gene expression (ex. [snRNA](https://en.wikipedia.org/wiki/Small_nuclear_RNA), [snoRNA](https://en.wikipedia.org/wiki/Small_nucleolar_RNA), [TERC](https://en.wikipedia.org/wiki/Telomerase_RNA), [lncRNA](https://en.wikipedia.org/wiki/Long_noncoding_RNA), [piRNA](https://en.wikipedia.org/wiki/Piwi-interacting_RNA), [siRNA](https://en.wikipedia.org/wiki/Small_interfering_RNA), and [shRNA](https://en.wikipedia.org/wiki/Short_hairpin_RNA))

**Table 3-2: Types of RNA polymerases found in eukaryotes and some of the RNAs they transcribe.**



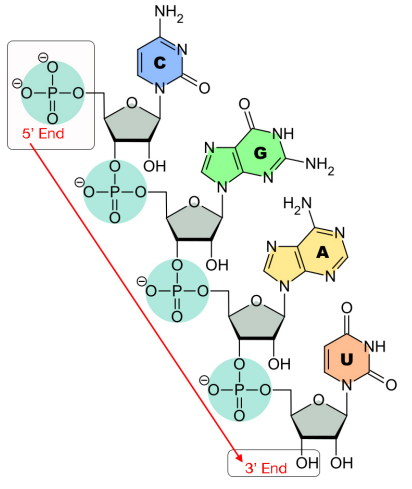
Building an RNA strand is very similar to building a DNA strand. This is not surprising, knowing that DNA and RNA are both nucleic acids, and as such are structurally very similar. Indeed, the three main differences between DNA and RNA are: 1) the presence of a 2’-OH group on the ribose sugar in RNA 2) uracil and not thymine is one of the pyrimidine nitrogenous bases in RNA and 3) DNA is double-stranded while the RNAs produced by transcription are single stranded. Transcription is catalyzed by the enzyme **RNA Polymerase**, which is a general term for an enzyme that makes a polymer of RNA. There are different RNA polymerases in eukaryotes that are responsible for synthesizing the many different RNAs found in cells (Table 3-2).



**Figure 3-1: *Bos taurus* RNA polymerase II.** The protein is shown in fuchsia; DNA is the double helix structure shown in the protein. The backbones of the DNA are orange and light pink. The RNA being synthesized cannot be seen well at this angle but is coming out of the back of the protein structure.

Like DNA polymerases, RNA polymerases build new RNA molecules in the 5' to 3' direction, but because they are making RNA, they use **ribonucleotides** (i.e., RNA nucleotides) rather than deoxyribonucleotides (as in DNA). Ribonucleotides are joined in exactly the same way as deoxyribonucleotides, which is to say that the 3'-OH of the last nucleotide on the strand is joined to the 5' phosphate of the incoming nucleotide.

A critical difference between DNA polymerases and RNA polymerases is that the latter do not require a primer to start making RNA. Once RNA polymerases are in the right place to start copying DNA, they just begin synthesizing RNA by stringing together RNA nucleotides that are complementary to the DNA template. The fact that RNA polymerases can copy DNA without a primer is known as *de novo* RNA synthesis, meaning RNA synthesis “starts from the beginning”. In contrast DNA replication requires a primer to initiate DNA synthesis (Chapter 14).



**Figure 3-2: RNA structure**. RNAs have a hydroxyl group at the 2’-carbon, while DNA does not. RNAs use the nitrogenous base uracil, unlike DNA, which uses thymine. Additionally, the strand is generated in the 5’ to 3’ direction (in the direction of the red arrow), when the incoming nucleotide is added, **its** 5’ phosphate is linked to the 3’OH of nucleotide that is already incorporated into the RNA strand.

## Transcription Initiation

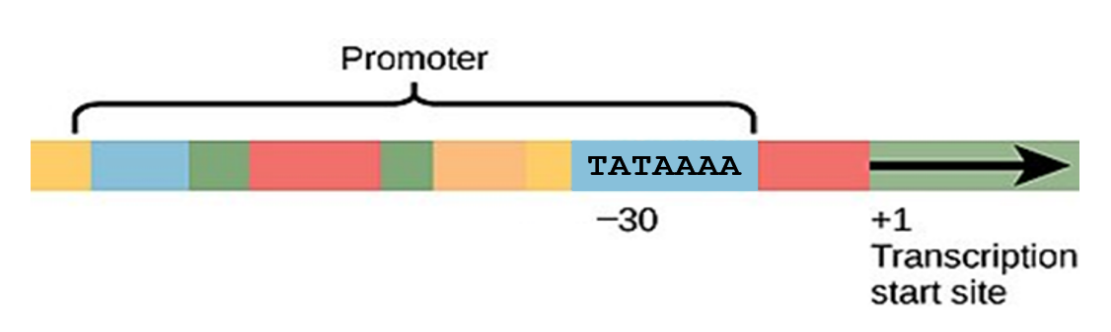
This, of course, brings us to an obvious question- how do RNA polymerases "know" where to start copying the DNA? Unlike the situation in replication, where every nucleotide of the parental DNA must eventually be copied, transcription only copies selected genes, coding portions of DNA, into RNA in a particular cell at any given time.

Consider the challenge: in a human cell, there are approximately 6 billion base pairs (bp) of DNA in the nucleus, of which about 2% stores protein-coding sequences. The small percentage of the genome that is made up of protein-coding sequences amounts to approximately 20,000 genes in each cell. Of these genes, only a small number will need to be expressed at any given time. For example, during S phase a cell will need helicase to promote DNA replication, but helicase is not needed during M phase, rather the cells need microtubules to complete M phase.

As shown in Table 3-1, RNA polymerase II is specific for transcription of mRNA. How does RNA polymerase II (RNA pol II) know where to start copying DNA to make an mRNA transcript? Signals in DNA direct RNA pol II to the location where it should start transcription. These signals are special sequences in DNA that are recognized by proteins (called **transcription factors**) that help RNA polymerase determine where it should bind the DNA to start transcription. A DNA sequence that signals the start of transcription is called a promoter. A promoter is often but not always located upstream of the gene that it controls. This means the promoter sequence is "before" the gene. In eukaryotes, promoters can be found in introns or in DNA sequences far away from the gene it regulates.

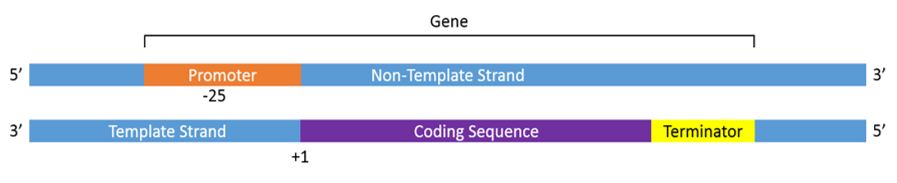
The promoter region for each gene has the ability to "control" the production transcription of the gene it is associated with. This is because expression of the gene is dependent on the binding of RNA polymerase to the promoter region to begin transcription. If the RNA polymerase and its helper proteins are unable to bind the promoter, the gene cannot be transcribed and it will therefore not be expressed. One common way to regulate gene expression is to prevent RNA polymerases and transcription factors from binding to the promoter sequence. When the DNA is tightly packaged, it does not allow the necessary proteins to access the DNA that is tucked away in a tight/condensed structure. (For additional information about how DNA packaging can affect gene transcription see Chapter 5)

What is special about a promoter sequence? To answer this question, scientists looked at many genes and their surrounding DNA sequences. This makes sense because RNA pol II has to bind to many different promoters to transcribe an array of DNA sequences. Therefore, the promoters should have some similarities in their sequences, so the RNA polymerase and transcription factors can easily determine where to start. Sure enough, common sequence patterns were seen to be present in many promoters. Some eukaryotic promoters contain a **TATA box**, a DNA sequence between 25 and 35 base pairs upstream of (before) the start of transcription (Figure 3-3 and 3-4). This sequence is recognized and bound by proteins that help RNA pol II position itself correctly to begin transcription. The consensus sequence for TATA boxes in eukaryotes is 5’-TATA(A/T)A(A/T)-3’. While not all promoters contain a TATA box, they all do contain some sort of signal sequence that helps to recruit proteins (transcription factors) that help RNA polymerases find where they are supposed to initiate transcription.



**Figure 3-3: Simplified structure of a eukaryotic promoter containing a TATA consensus sequence**. The +1 site denotes the start of transcription. Any sequences before the start site have a (-) designation. This is a very simplified model of a eukaryotic promoter where the TATA consensus sequence is between -25 and -35 nucleotides away from the start site. The other colors designate additional sequences found in the promoter that likely influence the regulation of this gene. Regulation of eukaryotic genes is often far more complex than depicted here.

The location where transcription begins is known as the +1 position or the **start site**. The start site is typically downstream of the promoter region. The first nucleotide to be transcribed is numbered +1. Any nucleotides before the +1 position have a negative number and any after have a positive number. In figure 3-4, notice that the TATA box is about 30 bases upstream of the +1 position, therefore it has a -30 designation. The TATA box and other DNA sequences of the promoter that are recognized by transcription factors are located on the **non-template** strand of DNA. This strand of DNA is not used as a template, and is complementary and antiparallel to the transcribed strand, called the **template strand**.

****

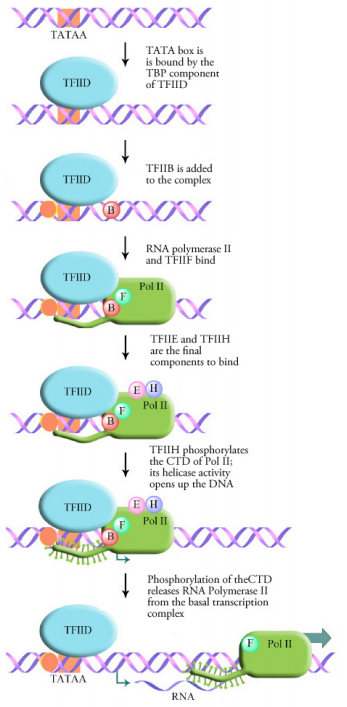
**Figure 3-4: Simplified structure of a gene.** The promoter is located on the non-template strand of DNA, which is oriented 5’ to 3’. The sequence that gets transcribed, the coding sequence and the terminator, are located on the template strand, which goes 3’ to 5’. Transcription of this gene would start at +1 and continue to the right.

The DNA for the promoter region is oriented in the 5’ to 3’ direction; thus the consensus sequences of the promoter are written in the 5’ to 3’ direction. The strand of DNA that is actually transcribed is called the template strand and is oriented 3’ to 5’; this is the strand that RNA polymerase II “reads” and uses to create RNA (Figure 3-4).

As noted previously, there are additional proteins called transcription factors that are required for RNA pol II to initiate transcription. What do these proteins do? **General transcription factors (GTFs)** are proteins that help eukaryotic RNA polymerases find transcription start sites and initiate RNA synthesis. We will focus on the transcription factors that assist RNA pol II. These transcription factors are named TFIIA, TFIIB and so on (TF= transcription factor, II=RNA pol II, and the letters distinguish individual transcription factors, but NOT the order they bind).

Transcription in eukaryotes requires the GTFs and RNA pol II to form a complex at the TATA box called the basal transcription complex or transcription initiation complex. The binding of these proteins is the minimum requirement for any gene to be transcribed. The first step in the formation of this complex is the binding of the TATA box by the transcription factor TFIID. TFIID has a binding site for the TATA box, called the **TATA binding protein** or TBP. When TBP binds to the TATA box, the DNA is bent causing the binding of additional transcription factors and eventually RNA pol II. After binding of TFIID, TFIIB is the second GTF to bind, and helps stabilize TFIID’s association and recruit RNA pol II and other associated GTFs. These steps are illustrated in figure 3-5.

The final step in the assembly of the basal transcription complex is the binding of a GTF called TFIIH. TFIIH is a multifunctional protein that has helicase activity (i.e., it is capable of opening up a DNA double helix) as well as kinase activity. The kinase activity of TFIIH adds phosphates onto the C-terminal domain (CTD) of the RNA pol II. This phosphorylation appears to be the signal that releases the RNA pol II from the basal transcription complex and allows it to move forward and begin transcription. In addition, phosphorylation of the CTD of RNA pol II is likely important for recruiting proteins that are important in mRNA modifications (capping, splicing and tailing-see below).



**Figure 3-5: Assembly of basal transcription complex and initiation of transcription.** TFIID binds the TATA box followed by TFIIB. This recruits RNA Polymerase II and TFIIF. TFIIE and TFIIH are the final GTFs to bind. TFIIH helps to activate transcription by phosphorylating the C-terminal domain (CTD) of RNA pol II and using its helicase activity to help unwind the DNA. Once RNA pol II has successfully created a short chain of RNA it will release from the basal transcription complex and begin the elongation phase of transcription.

## Transcription Elongation

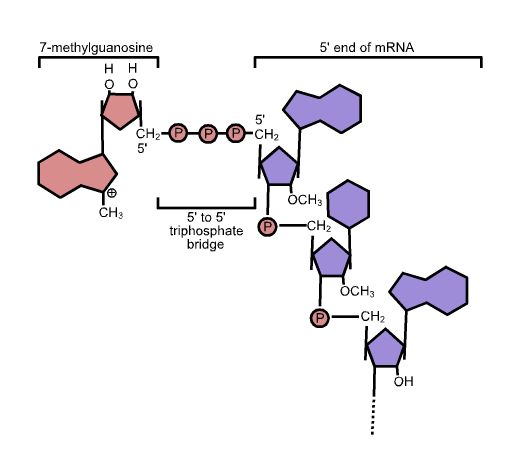
Sometimes RNA pol II may “fall off” of the DNA template (abortive initiation) before connecting enough RNA nucleotides together. This could be due to some instability in the binding of GTFs to the DNA, or other proteins within the transcription protein complex. Once the RNA pol II has synthesized enough RNA nucleotides to maintain a stable interaction with the DNA template, RNA pol II is released from a few transcription factors, such as TFIID and TFIIB, though TFIIF and TFIIH appear to remain bound. The release of these other proteins permits additional RNA polymerase II to initiate at the same promoter, thereby making additional mRNA copies of the same gene.

Within the active site of RNA pol II, the protein “reads” the DNA template in the 3’ to 5’ direction and binds the complementary ribonucleotide triphosphate, while generating a new strand of mRNA from 5’ to 3’ as is the norm for generation of nucleic acids. If the ribonucleotide that enters the active site base pairs correctly with the DNA base of the template, the enzyme will link the 5’ phosphate group of the incoming nucleotide with the 3’ hydroxyl group of the growing RNA chain. Because each ribonucleotide contains a high-energy triphosphate bond that is broken upon addition into the growing RNA chain, no additional energy is required to build a polymer of RNA. Once the new ribonucleotide is connected with the previous ribonucleotide, the RNA polymerase will translocate, move down one nucleotide and repeat this action until the RNA polymerase is eventually removed from the template during transcription termination. Because RNA pol II repeats this action of reading a DNA base, adding a ribonucleotide over and over without stopping, typically, the enzyme is considered to be processive (this can be contrasted with an enzyme that does a reaction once or a few times and then ‘falls off’).

**Adding a 5’-cap to the mRNA Transcript**

The initial product of transcription of a protein coding gene is called the **pre-mRNA** (or primary transcript). After it has been processed and is ready to be exported from the nucleus, it is called the mature mRNA or processed mRNA. These modifications include the addition of a 5’ cap, a 3’ poly-A tail and the removal/splicing of introns.

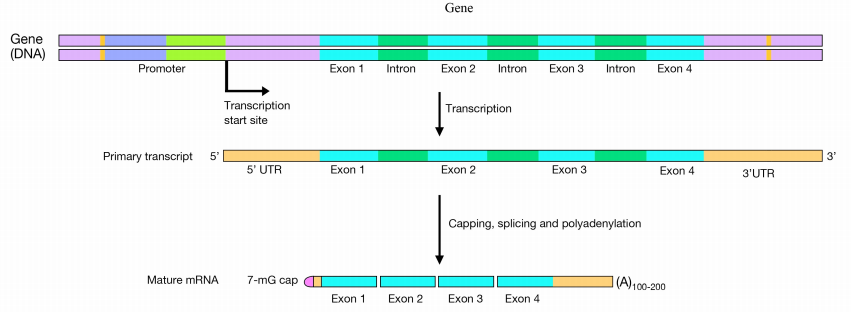
Once the new mRNA transcript emerges from the RNA polymerase, there is a capping enzyme associated with the phosphorylated CTD of RNA poly II. This enzyme is responsible for capping the 5’ end of the mRNA during the process of transcription. In the capping step of mRNA processing, a 7-methyl guanosine (Fig 3-6) is added at the 5' end of the mRNA. The 5’ end of the mRNA is the first part of the transcript that is transcribed. The **5’ G cap** protects the 5' end of the mRNA from degradation by nucleases, allows for its proper transport out of the nucleus, and also helps to position the mRNA correctly on the ribosomes during protein synthesis. As depicted in figure 3-15, this process occurs in the nucleus of eukaryotic cells.



**Figure 3-6: The 5’ cap.** The modified nucleotide, 7-methylguanosine, is added to the 5’ end of the growing mRNA transcript.

**mRNA Splicing**

Most eukaryotic genes and theirpre-mRNAtranscripts contain noncoding stretches of nucleotides or regions that will not be made into protein. These noncoding segments are called**introns** and must be removed before the mature mRNA can be transported to the cytoplasm and translated into protein.  The stretches of DNA that code for amino acids in the protein are called **exons.**  During the process of splicing, introns are removed from the pre-mRNA by the spliceosome, and exons are spliced back together.  If the introns are not removed, the RNA would be translated into a nonfunctional protein.  Splicing occurs in the nucleus before the RNA migrates to the cytoplasm, and happens during the process of transcription elongation.  Once splicing is complete, the mature mRNA (containing uninterrupted coding information), is transported to the cytoplasm where ribosomes translate the mRNA into protein. Figure 3-7 demonstrates the presence of introns and exons in the DNA that are transcribed into RNA, where the introns are removed before the mature mRNA can be exported from the nucleus.



**Figure 3-7: Splicing.** Introns and exons are encoded in the DNA (top) and transcribed into the primary transcript (middle).The introns are removed via the process of splicing, resulting in a mature mRNA where all introns have been removed (bottom).

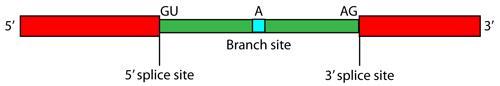
There are two main steps in splicing. In the first step, the pre-mRNA is cut at the 5' splice site (the junction of the 5' exon and the intron). The 5' end of the intron then is joined to what is called the **branch point**, a specific location within the intron. This generates a “lasso” shaped structure called the **lariat**, a characteristic of the splicing process. In the second step, the 3' splice site is cut, the two exons are joined together, and the intron is released.

We will now go through the process of splicing in more detail.  In this example, the pre-mRNA contains two exons and one intron (Fig 3-8).

A primary mRNA transcript containing two exons (red) and one intron (green). 

**Figure 3-8: Splicing part 1, the transcript.** A primary mRNA transcript containing two exons (red) and one intron (green).

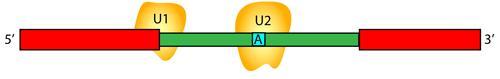
As mentioned, introns contain several conserved sequences that guide the splicing process: a 5’ GU sequence called the 5’ splice site (GU) , a branch site (A) and a 3’ splice site (AG), as shown in figure 3-9.



**Figure 3-9: Splicing part 2, conserved sequences.** A primary mRNA transcript containing two exons (red) and one intron (green), demonstrating the location of the 5’ (GU) and 3’ (AG) splice sites located on each end of the intron, and the branch site (A) located within the intron (blue).

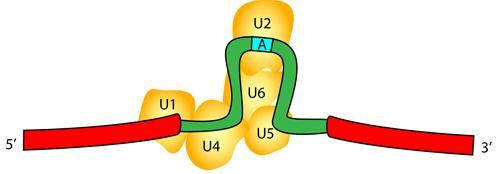
**The Spliceosome**

A large complex known as the spliceosome controls RNA splicing. The spliceosome is composed of particles made up of both RNA and protein.  These particles are called **small nuclear ribonucleoprotein**or**snRNPs** (pronounced “snurps”) for short.  The snRNPs recognize the conserved sequences within introns and quickly bind these sequences once the pre-mRNA is made and initiate splicing. The spliceosome is built in distinct steps.  First, specific snRNPs (U1 and U2) bind the 5’ splice site and the branch site (Fig 3-10)**.**



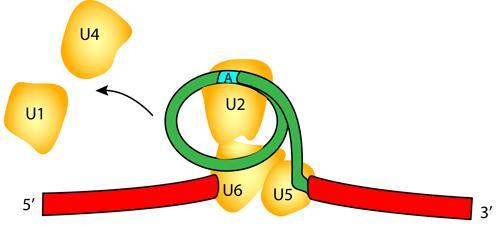
**Figure 3-10: Splicing part 3, initial snRNP binding.** A primary mRNA transcript containing two exons (red) and one intron (green), demonstrating the binding of specific snRNPs (labeled U1 and U2) to the mRNA at the 5’ splice site and the branch site.

Several other snRNPs (U4-6) then bind the pre-mRNA transcript forming the mature spliceosome complex**.**  This causes the intron to form a loop and brings the 5’ splice site and 3’ splice site together (Fig 3-11).



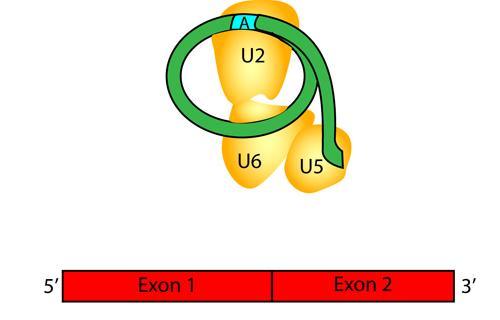
**Figure 3-11: Splicing part 4, the mature spliceosome complex.** A primary mRNA transcript containing two exons (red) and one intron (green), demonstrating the location of snRNP (U1, U2, U4, U5, and U6) binding to form the mature spliceosome complex.

Now that the spliceosome is assembled, the process of splicing can begin.  First, the 5’ end of the intron is cut.  The 5’ (GU) end of the intron is then connected to the A branch site, which creates a lariat structure (Fig 3-12).



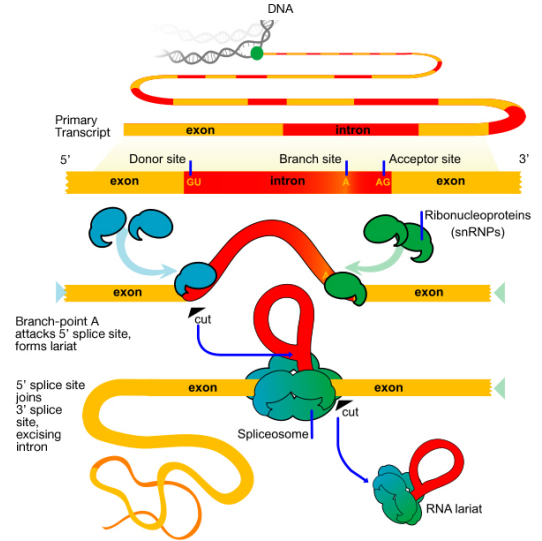
**Figure 3-12: Splicing part 5, lariat structure formation.** A primary mRNA transcript containing two exons (red) and one intron (green), demonstrating the generation of a lariat (“lasso”) structure via snRNPs, where U1 and U4 are then released.

At this stage, the 3’ splice site is cleaved.  Once the intron has been fully cleaved, the two exons are covalently attached.  The intron in the form of a lariat is released along with the rest of the snRNPs (Fig 3-13).



**Figure 3-13: Splicing part 6, release of the lariat.** Aprimary mRNA transcript containing two exons (red) covalently joined together and the release of the intron and the remaining components of the spliceosome.

The spliced intron will then be degraded and the snRNPs are used again to splice other pre-mRNAs.  This process occurs at multiple locations within the mRNA, and all introns must be removed for the transcript to be exported and ready for translation within the cytoplasm. Fig 3-14 combines all of the steps outlined above in one summary.



**Figure 3-14: Splicing of pre-mRNA.** The entire splicing process, from a DNA molecule containing introns and exons, to a fully spliced transcript wherethe spliceosome removes introns as lariat structures.

## Transcription Termination and Polyadenylation

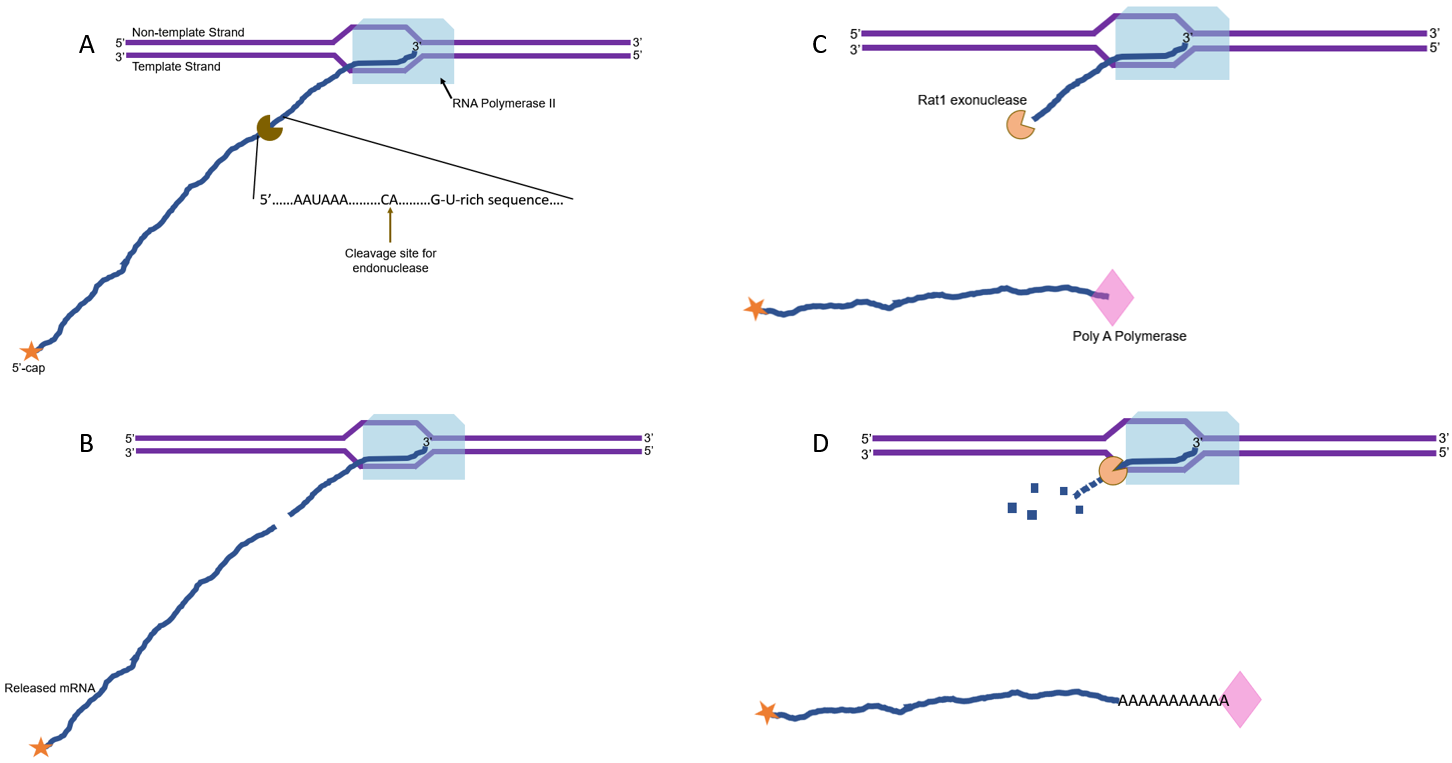
Termination of transcription is not as well understood in eukaryotes as it is in prokaryotes, but we do know that termination is coupled with polyadenylation of the 3’ end of the transcript.

**Polyadenylation**

Recall that the CTD of RNA pol II is phosphorylated; it is thought that phosphorylation at specific locations in the CTD recruits proteins that cleave the mRNA at a specific sequence. While RNA Polymerase II is still transcribing, the pre-mRNA is cleaved by an endonuclease-containing protein complex between an AAUAAA consensus sequence, and a GU-rich sequence. This releases the functional mRNA from the rest of the transcript, which is still attached to the RNA Polymerase. An enzyme called poly (A) polymerase (PAP) is part of the same protein complex that cleaves the pre-mRNA and it immediately adds a string of approximately 200 A nucleotides, called the **poly (A) tail**, to the 3′ end of the just-cleaved pre-mRNA. The poly (A) tail protects the mRNA from degradation, aids in the export of the mature mRNA to the cytoplasm, and is involved in binding proteins involved in initiating translation. As depicted in Figure 3-15, similar to the addition of the 5’ cap, polyadenylation also occurs in the nucleus of eukaryotic cells.

**Termination**

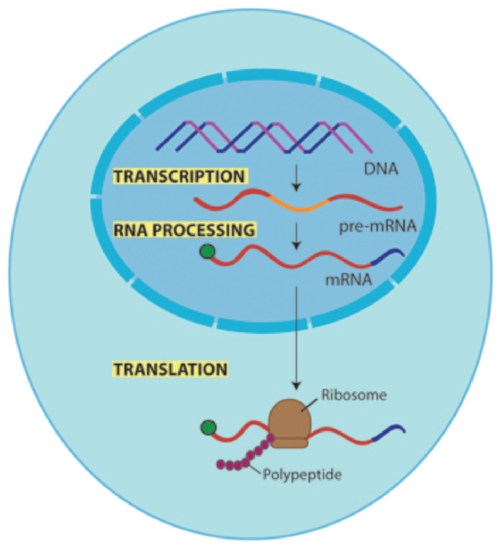
Once the mRNA is released, the RNA pol II continues to transcribe DNA. Termination occurs when the RNA pol II interaction with the DNA template is weakened, and the RNA polymerase unbinds from the DNA. It is unclear how exactly this occurs, but it is thought that an exonuclease (called Rat1) binds to the newly cleaved end of the RNA that protrudes from the RNA pol II. The exonuclease ‘chews’ up the RNA as it moves towards the RNA polymerase. This protein may catch-up to the RNA pol II and cause it to pause; this pause may allow another protein that has a helicase function to bind and unwind the DNA/RNA interaction found near the active site of the RNA polymerase. This may cause the RNA polymerase’s interaction with the DNA template to become very weak, and the protein may ‘fall off’ of the template, terminating transcription (Figure 3-15)



**Figure 3-15: One model of termination in eukaryotes.** (A) A cleavage site is recognized in the mRNA and (B) cleaved by an endonuclease. This releases the mRNA. (C) Rat1 exonuclease starts cleaving the 5’ end of the RNA that is still attached to RNA polymerase II and Poly A polymerase (PAP) binds to the 3’ end of the mRNA. (D) Rat1 continues ‘chewing’ the RNA until it physically hits RNA pol II and destabilizes the interaction of pol II with the DNA and it eventually ‘falls off’. Additionally, PAP adds many adenine nucleotides to the 3’ end of the released mRNA. Note at this point the mRNA has a 5’ cap, has been spliced and once polyadenylated, it will be ready for export out of the nucleus.

# Location of Transcription and Processing in Eukaryotes

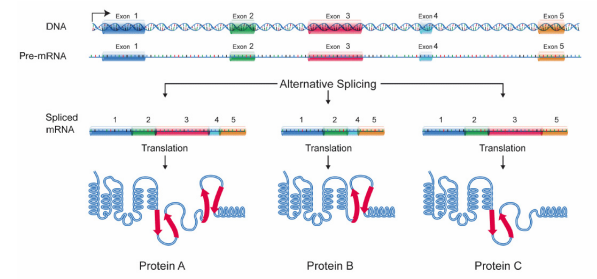
Transcription and processing (which includes splicing) of the newly made mRNA occurs in the nucleus of the cell.  Once a mature mRNA transcript is made it is transported to the cytoplasm for translation into protein (Fig 3-16).



**Figure 3-16: Post-transcriptional RNA modifications in a eukaryotic cell**. RNA processing events including addition of the 5’ cap, splicing and polyadenylation occur in the nucleus. The mature mRNA, or fully processed mRNA transcript, is exported from the nucleus to the cytoplasm for translation.

## Alternative Splicing

Many pre-mRNAs have a large number of exons that can be spliced together in different combinations to generate different mature mRNAs. This is called **alternative splicing** and allows the production of many different proteins using relatively few genes. This is true because a single RNA can, by combining different exons during splicing, create many different protein-coding messages. Because of alternative splicing, each gene in our DNA gives rise to, on average, to three different proteins, therefore there can be many more proteins in our proteome, than genes in our genome. Figure 3-17 demonstrates how one DNA molecule that contains five separate exons and four introns can be spliced to produce three different mRNAs. These mRNAs would then be exported to the cytoplasm for translation and, therefore, produce three different protein products. The spliced mRNA on the left shows splicing that removed each intron while connecting the existing exons. Alternatively, the spliced mRNA in the middle and on the right have had entire exons removed during the splicing process, exon 3 and exon 4, respectively. Note that when mRNAs are alternatively spliced, there can be no rearrangement of exons, only the removal of introns and, depending on the splice variant, individual interior exons can also be spliced out. As can be seem in Figure 3-17 several different proteins can be generated from one gene, demonstrating that there can be more protein products than the number of genes present in eukaryotic cells.



**Figure 3-17: Alternative splicing.** Alternative splicing allows for multiple proteins to be generated from a single region of DNA. In the example above, one pre-mRNA, containing five exons and four introns, generates three distinct protein products.

## Mature mRNA

After all RNA modifications have been completed, including the addition of the 5’ cap, the poly-A tail and the splicing of introns, the mature mRNA is ready for export from the nucleus into the cytoplasm. Figure 3-17 represents a fully processed mRNA that is ready for translation. Note that the mature mRNA contains the 5’ cap, the poly A tail, and a coding region made up only of exons. Additionally, portions of the mRNA at the 5’ end and 3’ end of the transcript will not be translated, called **UTR’s** (untranslated regions). The 5’ UTR includes everything in the mRNA transcript before the start codon, including the 5’ cap, while the 3’ UTR includes everything between the stop codon and the end of the transcript, including the poly-A tail.



**Figure 3-18: A fully processed mRNA transcript**. After all RNA modifications have been completed, a mature mRNA contains a 5’ cap, a poly-A tail, and untranslated regions (UTRs) at the 5’ and 3’ ends.

Associated videos:

1. [RNA processing (NDSU Virtual Cell Animations Project animation 'mRNA Processing')](https://www.youtube.com/watch?v=YjWuVrzvZYA).

[More information.](http://vcell.ndsu.edu/animations)