SECTION 3. RNA TRANSCRIPTION AND POST – TRANSCRIPTIONAL PROCESSING

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain how RNA is synthesized using DNA as a template
- Distinguish between transcription in prokaryotes and eukaryotes
- Describe post -transcriptional processes that eukaryotic pre-mRNA undergo before translation

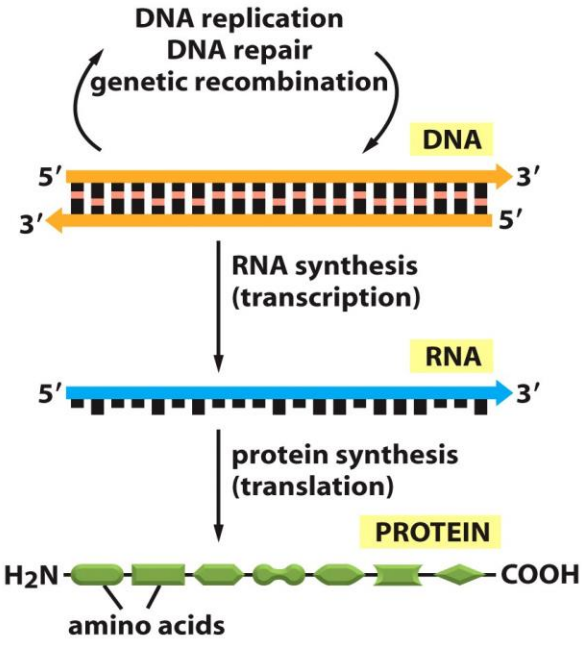


Figure 1. The central dogma of molecular biology. *Molecular Biology of the Cell* (© Garland Science 2008)

The genes present in an organism's genome will not have any effect on its functions until they are "expressed". Gene expression refers to the production of a protein or functional RNA from its gene.

The information flow and utilization of the genome of an organism is a two – step process: DNA is transcribed to RNA via complementary base pairing, and the RNA transcript is then translated to a protein product. Final folding and modifications produce the functional form of the proteins that actually do things in cells. This series of processes is known as the **central dogma of molecular biology**, and holds true for all living things.

The gene is the functional unit of heredity. It consists of two essential regions:

1. transcriptional region – the DNA segment that will be transcribed into mRNA
2. regulatory region – not transcribed, but contains sequences that are recognized by the transcriptional machinery

Traditionally, genes were only thought to produce one gene product – proteins. However, recent advances in genomics have shown that many genes in fact produce non – protein coding RNA as their gene product.

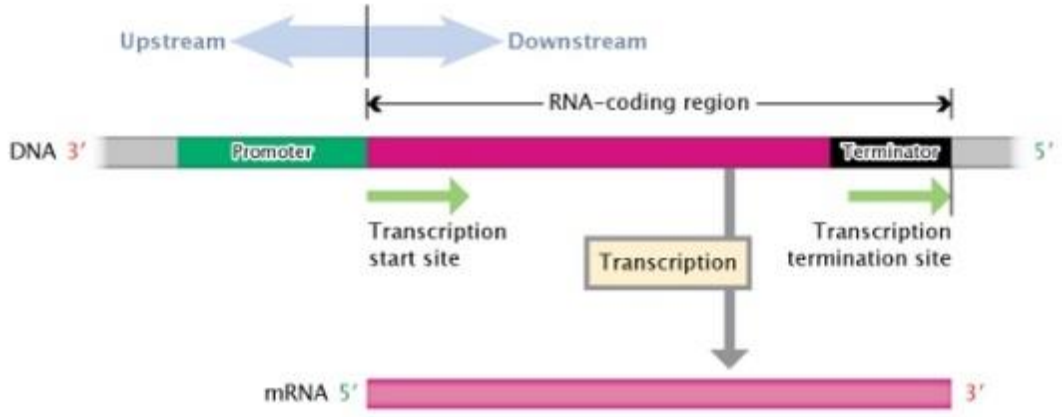


Figure 2. A DNA transcription unit. A DNA transcription unit is composed, from its 3' to 5' end, of an RNA-coding region (pink rectangle) flanked by a promoter region (green rectangle) and a terminator region (black rectangle). Regions to the left, or moving towards the 3' end, of the transcription start site are considered "upstream", while regions to the right, or moving towards the 5' end, of the transcription start site are considered "downstream." © 2014 **Nature Education** Adapted from Pierce, Benjamin. *Genetics: A Conceptual Approach*, 2nd ed

Transcription in both prokaryotes and eukaryotes takes place in three steps: initiation, elongation, and termination. However, there is significantly more complexity in eukaryotic transcription. The presence of organelles in eukaryotic cells means that transcription needs to finish first before translation can occur, whereas in prokaryotes, RNA that is still being synthesized may already start to be translated.

In addition, the transcribed region of eukaryotic genes contains both **exons**, which correspond to protein-coding sequences, and intervening sequences called **introns** which do not encode functional proteins. Introns need to be removed, or spliced out, before translation can occur.

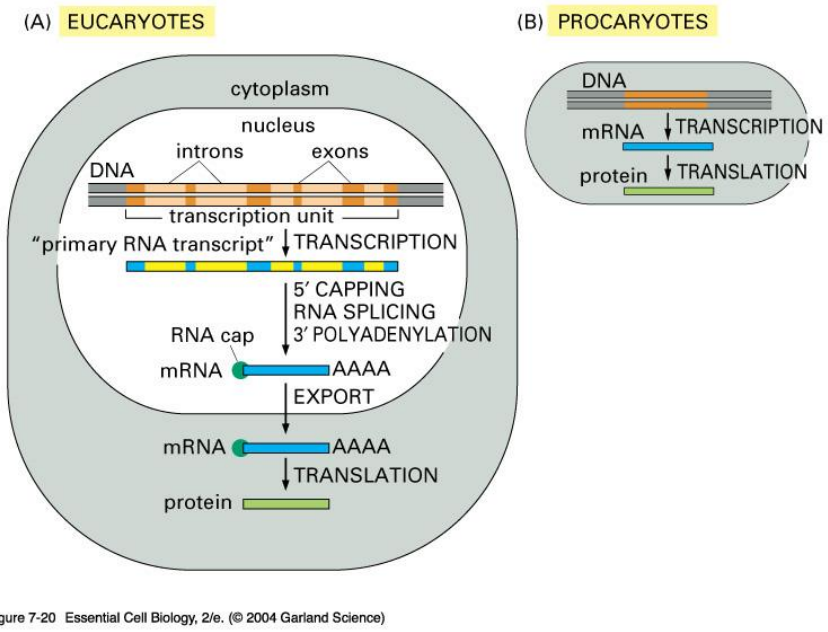


Figure 7-20 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Figure 3. Differences in the gene expression processes between prokaryotes and eukaryotes arise from the higher complexity of eukaryotic cells and genomes.

PROKARYOTIC TRANSCRIPTION

Since prokaryotes lack membrane-bound nuclei, transcription occurs in the cell cytoplasm. Transcription products also often produce more than one gene product. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are transcribed as polycistronic mRNAs from DNA transcription units known as **operons**.

The **promoter** is a DNA sequence from which the transcription machinery binds and initiates transcription. Two consensus sequences regions that are similar across all promoters are found to be present across various bacterial species: the -10 and -35 regions upstream of the transcription initiation site (indicated as +1). These sequences are recognized by RNA polymerase, and are critical in the initiation of transcription. Deviation from the consensus sequence usually leads to weaker gene expression.

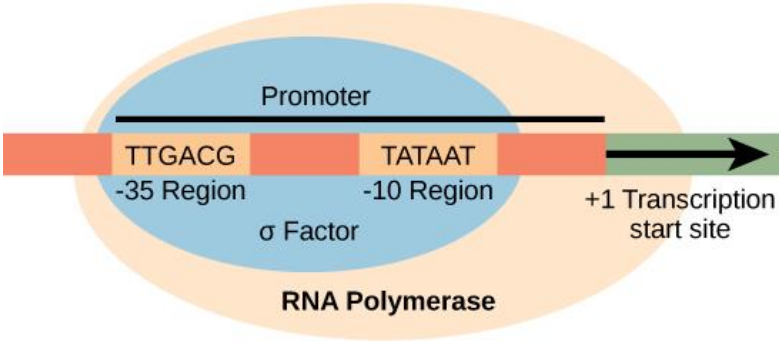


Figure 3. The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start sight. The σ subunit dissociates from the polymerase after transcription has been initiated. From https://courses.lumenlearning.com/wm-biology1/chapter/prokaryotic-transcription/

RNA polymerase is a multi-subunit protein composed of the following polypeptide subunits:

- σ - confers transcriptional specificity; involved only in transcription initiation
- 2 units α – for assembling the polymerase on the promoter
- *β* - binds to the NTP that will become part of the newly synthesized mRNA molecule
- *β'* - binds the DNA template strand

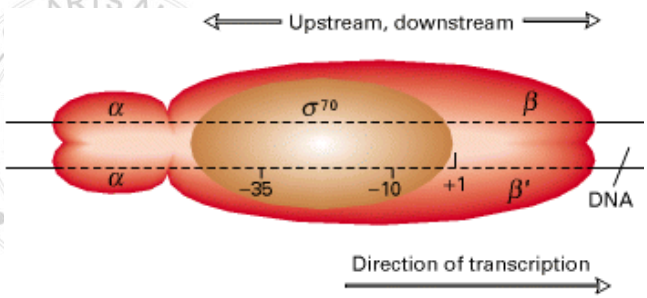


Figure 4. Positioning of the RNA polymerase subunits in the prokaryotic promoter region.

The RNA polymerase holoenzyme assembles every time a gene is transcribed, and disassemble once transcription is complete. The -35 sequence is recognized and bound by *σ* subunit, which acts as an "initiation factor". It is then joined by the other subunits join to form the RNA polymerase holoenzyme. The elongation phase begins with the release of *σ* from the holoenzyme. Its dissociation allows the remaining subunits (the core RNA polymerase) to disrupt H-bonds between the double stranded template.

Both RNA and DNA polymerases can add nucleotides to an existing strand, extending its length. However, unlike DNA polymerase, RNA polymerase can initiate the synthesis of a new strand. Therefore, no primer is needed for transcription. RNA polymerase synthesizes mRNA in the 5' to 3' direction, and the unit of addition is a nucleotide triphosphate (NTP). As in DNA replication, two phosphate groups are released as pyrophosphate (PP_i) during the reaction. The first nucleotide at the 5' end retains its triphosphate group. Transcription produces a single-stranded RNA molecule that is complementary to the DNA template.

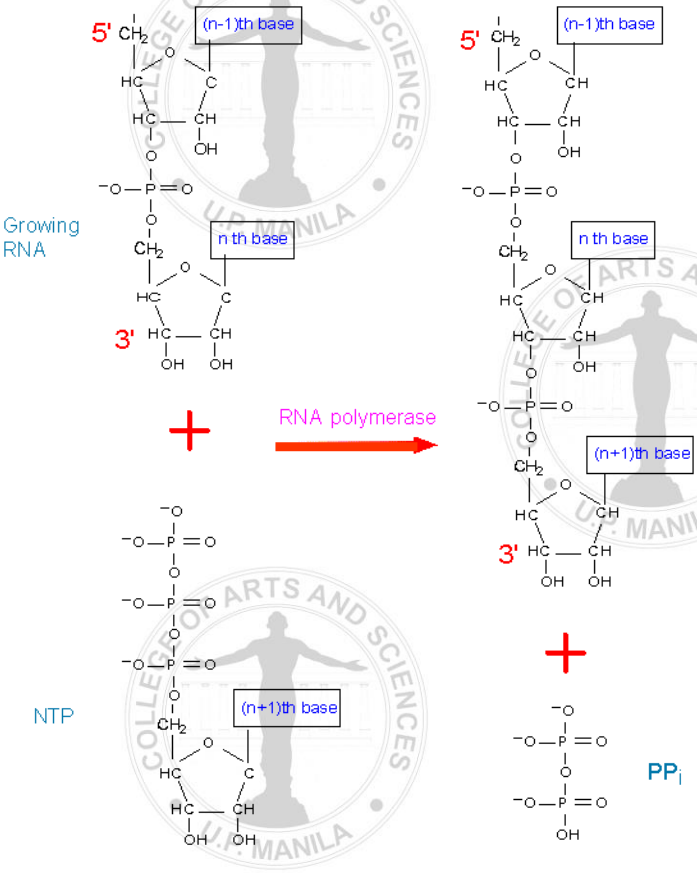


Figure 5. The chemical reaction catalyzed by RNA polymerase. From https://www.web-books.com/MoBio/Free/Ch4B1.htm

As elongation proceeds, a transcription bubble is formed, containing a hybrid between the template DNA strand and the nascent RNA, while the non – template strand is temporarily single stranded. DNA ahead of the transcription bubble is overwound, while DNA behind the transcription bubble is underwound. Topoisomerases relieve these torsional stresses.

Unwinding of the double strand by RNA polymerase produces a transcription bubble consisting of ~25 unwound DNA basepairs. About 8 nucleotides of newly-synthesized RNA transiently basepairs with the template DNA. The rest of the RNA molecules falls off the template, allowing the DNA behind it to rewind.

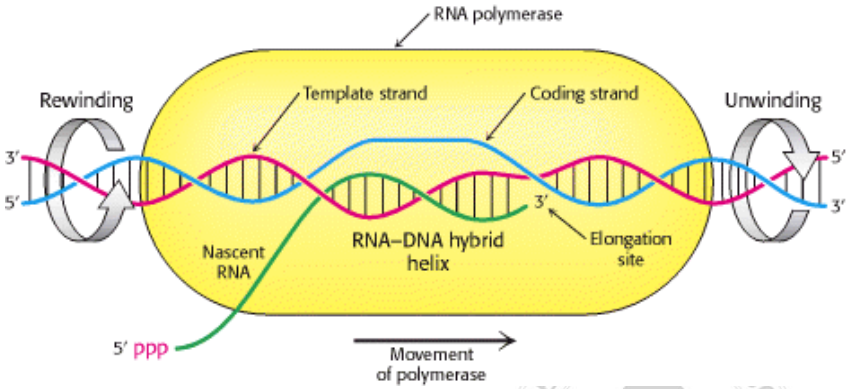


Figure 5. The transcription bubble features transiently separated DNA double strand and a hybrid formed between the DNA template and the nascent

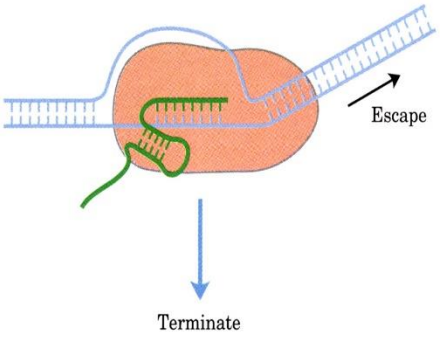
Note that only one of the two DNA strands will be used as a template to form the RNA transcript. The non – template strand is also called the coding strand, since it has the same sequence as the RNA transcript, except for the presence of U instead of T in the RNA.

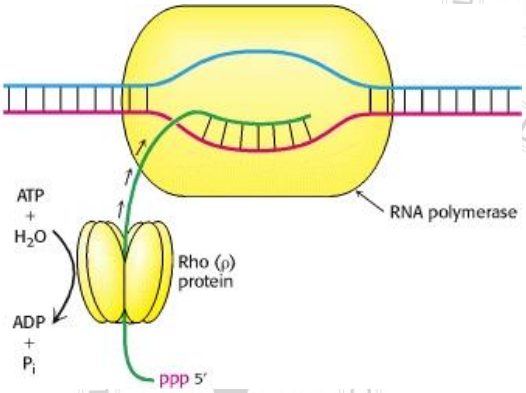
```
5' ACATCGACGCGCAGTTAATCCC..3' DNA coding strand (+)
3' TGTAGCTGCGCGTCAATTAGGG..5' DNA template strand (-)
5' ACAUCGACGCGCAGUUAAUCCC..3' RNA (+)
```

Prokaryotic termination of transcription

In bacteria, there are two known termination mechanisms:

1. **intrinsic termination:** sequences within the growing RNA chain, specifically, a GC-rich region, forms a stem-loop structure at the 3' end of the RNA, which weakens the RNA-DNA hybrid and allows RNA to dissociate from the transcription bubble
2. **rho-dependent termination:** rho-factor protein binds to the growing RNA transcript at a C-rich region and pulls RNA chain away from RNA polymerase and the DNA template. As a result, the rho protein collides with the polymerase, and the mRNA is released from the transcription bubble.





Even before termination of transcription, the prokaryotic transcript may already be used to synthesize protein. The lack of compartmentalization in the prokaryotic cell means these processes can occur concurrently.

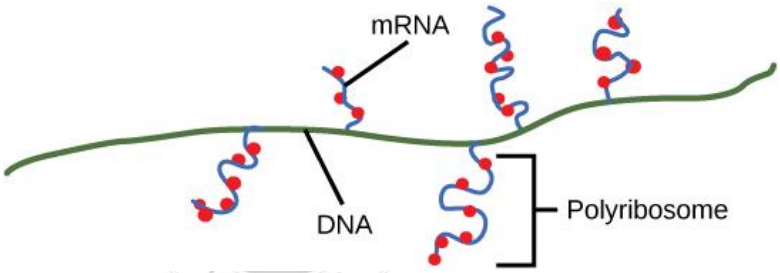


Figure 6. Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

From https://courses.lumenlearning.com/wm-biology1/chapter/prokaryotic-transcription/

Transcription in Eukaryotes

Because eukaryotic cells are more complex than prokaryotic cells, transcription is also more complex. Instead of a single multi-subunit RNA polymerase, eukaryotes have three multi-subunit that transcribe differents set of genes. RNA Polymerase II is the enzyme used to synthesize protein – coding mRNA. The three eukaryotic RNA polymerases can be differentiated based on their sensitivity to α-amanitin, a poisonous compound from *Amanita phalloidesi.* RNA polymerase II is extremely sensitive to it, whereas RNA polymerase I completely insensitive is and RNA polymerase III is moderately sensitive to the compound. Table 1 describes the three eukaryotic DNA polymerases.

Table 1. Locations, Products, and Sensitivities of the Three Eukaryotic RNA Polymerases

RNA Polymerase	Cellular Compartment	Product of Transcription	α-Amanitin Sensitivity
I	Nucleolus	All rRNAs except 5S rRNA	Insensitive
II	Nucleus	All protein-coding nuclear pre-mRNAs	Extremely sensitive
III	Nucleus	5S rRNA, tRNAs, and small nuclear RNAs	Moderately sensitive

https://courses.lumenlearning.com/wm-biology1/chapter/reading-rna-polymerase/

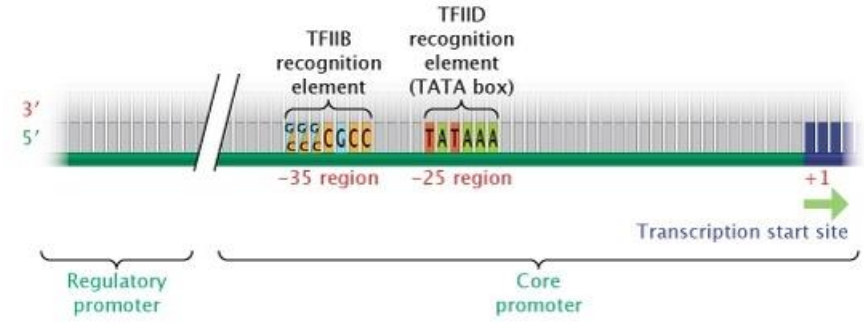


Figure 7: Eukaryotic core promoter region. In eukaryotes, genes transcribed into RNA transcripts by the enzyme RNA polymerase II are controlled by a core promoter. A core promoter consists of a transcription start site, a TATA box (at the -25 region), and a TFIIB recognition element (at the -35 region).

Transcription of protein – coding genes requires RNA Polymerase II – associated transcription factors (TFIIs). One of these, TFIID, contains the TATA-binding protein (TBP) which recognizes the TATA box and ensures that the correct start site is used. Binding of TFIID recruits other transcription factors, including TFIIB, TFIIE, TFIIF, and TFIIH to the TATA box. Once this complex is assembled, RNA polymerase can bind to its upstream sequence to complete the formation of the pre-initiation complex.

Following the formation of the preinitiation complex, RNA polymerase is released from the other transcription factors, and elongation is allowed to proceed as it does in prokaryotes, with the polymerase synthesizing pre-mRNA in the 5' to 3' direction.

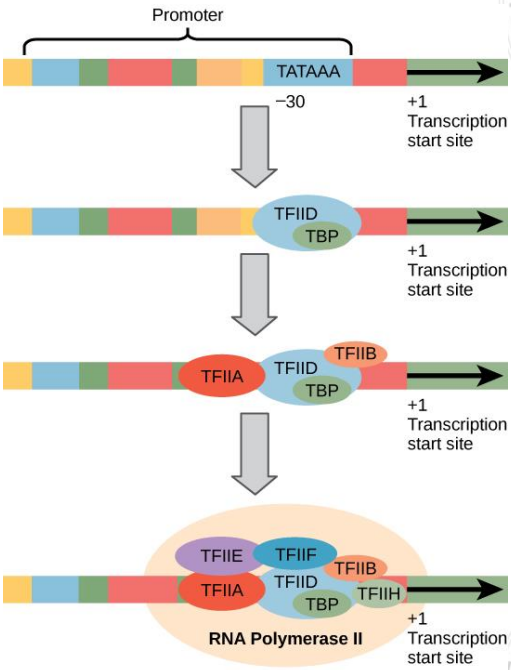


Figure 8. A generalized promoter of a gene transcribed by RNA polymerase II is shown. Transcription factors recognize the promoter. RNA polymerase II then binds and forms the transcription initiation complex. https://courses.lumenlearning.com/wm-biology1/chapter/reading-rna-polymerase/

As in prokaryotic transcription, unwinding of the double strand by RNA polymerase produces a transcription bubble of about 25 transiently separated DNA basepairs. RNA polymerase travels along the template DNA and adds nucleotides one a time until termination signals are reached.

Transcription termination in eukaryotes is not that well – understood, but it is known to involve a **polyadenylation signal** that appears in the RNA transcript. The polyadenylation signal is recognized by an enzyme that cuts the RNA transcript downstream of that sequence, releasing it from RNA polymerase.

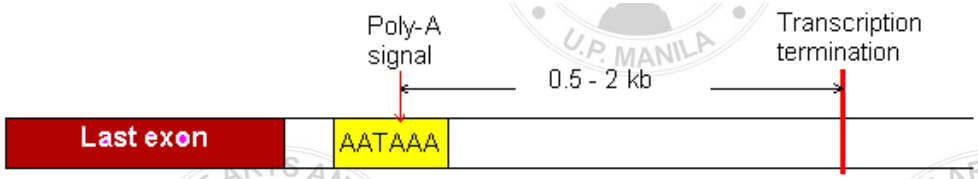


Figure 9. RNA polymerase terminates transcription downstream of a poly-A sequence. From https://www.web-books.com/MoBio/Free/Ch4E.htm

Post – transcriptional processing

The immediate product of eukaryotic transcription is not yet mature messenger mRNA. Eukaryotic pre-mRNA must undergo extensive processing, specifically, the addition of a 5' cap,

polyadenylation, and the removal of introns (splicing) before it can be transported out of the nucleus and translated to protein.

1. **Addition of 5’ cap**

   Even while the pre-mRNA is still being synthesized, a **7-methylguanosine cap** is added to the 5' end of the growing transcript by a phosphate linkage. This functional group protects the nascent mRNA from degradation. The 5’ cap also plays an important role in the initiation of translation by ribosomes.

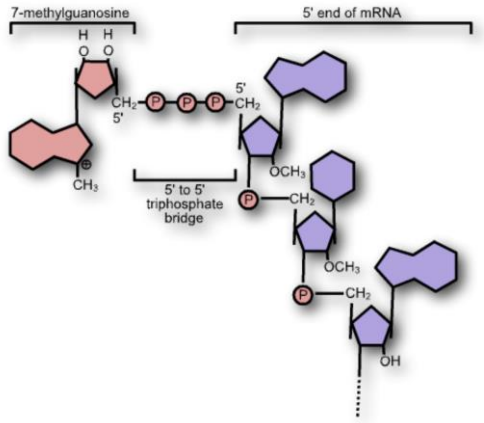


Figure 10. The 5’ cap, placed on the 5’ end of the mRNA, is a methylated guanosine triphosphate molecule (GTP) linked to the first RNA nucleotide by a special 5'-5' triphosphate linkage. From https://courses.lumenlearning.com/boundless-biology/chapter/rna-processing-in-eukaryotes/

2. **Addition of a 3’ poly(A) Tail**

   Poly-A polymerase enzyme adds a poly(A) tail consisting of ~200 residues 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.

3. **Pre-mRNA Splicing**

   All of a pre-mRNA’s introns must be completely and precisely removed, or spliced out, before the pre-mRNA leaves the nucleus. If the process errs by even a single nucleotide, the resulting protein may end up being dysfunctional.

   Splicing is usually carried out the spliceosome complex, which consists of small nuclear ribonucleoparticles (snRNPs) that are themselves made up of a complex of proteins and small nuclear RNAs (snRNAs). The spliceosome cleaves the 5’-GU end of the intron and then covalently attaches the G to an internal A nucleotide in the branch site. It then connects the 3’ end of the first exon to the 5’ end of the next exon, cleaving the 3’ end of the intron in the process. This results in the splicing together of the two exons. The intron is released in lariat form.

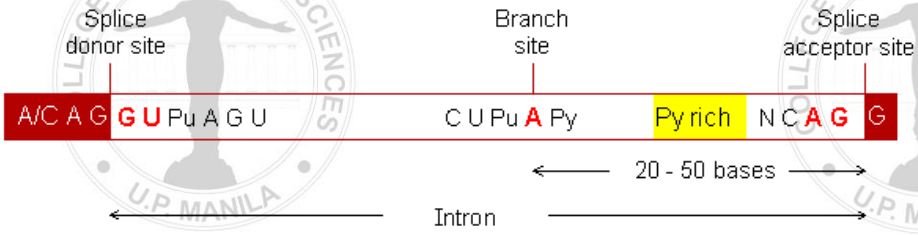


Figure 11. The consensus sequence in introns that are used for splicing. Pu = A or G; Py = C or U. From https://www.webbooks.com/MoBio/Free/Ch5A4.htm

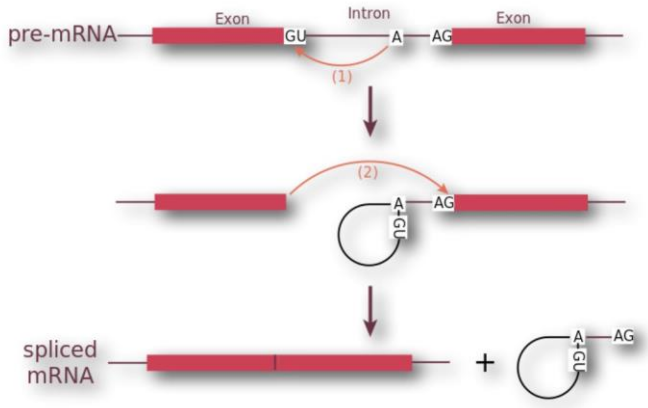


Figure 12. Pre-mRNA splicing involves the precise removal of introns from the primary RNA transcript. Spliceosomes recognize sequences at the 5' and 3' end of the intron. Initially, the conserved G which starts an intron is cleaved from the 3' end of the exon upstream to it and the G is covalently attached to an internal A within the intron. Then the 3' end of the just-released exon is joined to the 5' end of the next exon, cleaving the bond that attaches the 3' end of the intron to its adjacent exon. This both joins the two exons and removes the intron in lariat form.

Figure 13 summarizes the post – transcriptional processes undergone by pre-mRNA.

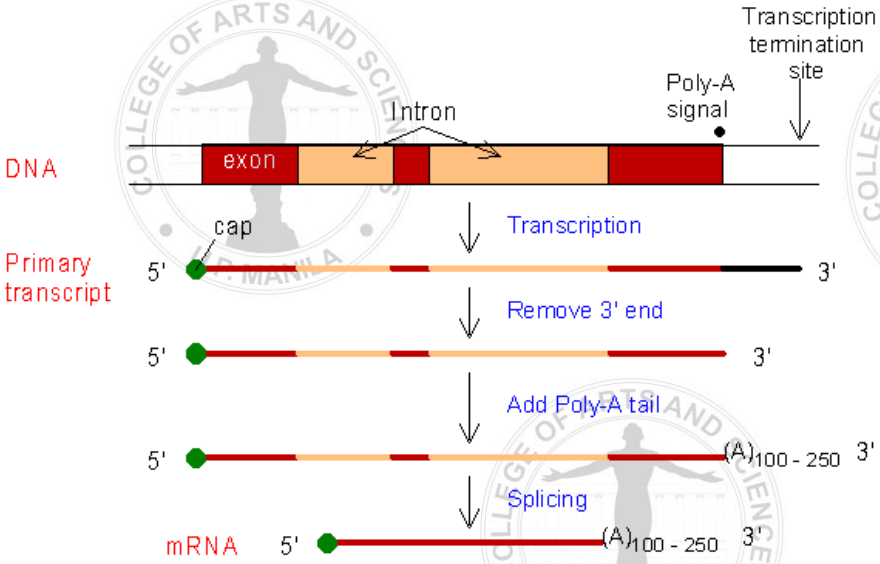


Figure 13. The procedure of pre-mRNA processing for protein coding genes. From https://www.web-books.com/MoBio/Free/Chap5.htm

Regulation of eukaryotic transcription

All organisms and cells control or regulate the transcription and translation of their DNA into protein. Controlling eukaryotic gene expression is important because it allows cells to

- express genes when needed
- repress genes when not needed
- conserve energy resources by avoiding expressing unnecessary/detrimental genes

The regulation of gene expression can occur at all stages of the process of gene expression. However, the easiest and most common stage of control is at the transcription level.

Eukaryotic gene regulation transcription is a much more complex process compared to that in prokaryotes. Aside from RNA polymerase and its associated transcription factors that assemble to form the pre-initiation complex, other DNA sequences and associated proteins interact to regulate the frequency with which a gene is transcribed to mRNA.

Eukaryotic genes usually contain enhancers, which are DNA regions that are recognized by activator proteins that facilitate, or enhance, transcription. Enhancers can be located upstream or downstream of a gene, and may be thousands of nucleotides away. Binding of the activator protein to the enhancer leads to a conformational change in the DNA that leads to the active recruitment of the general transcription factors and RNA polymerase, and thus, faster assembly of the pre-initiation complex.

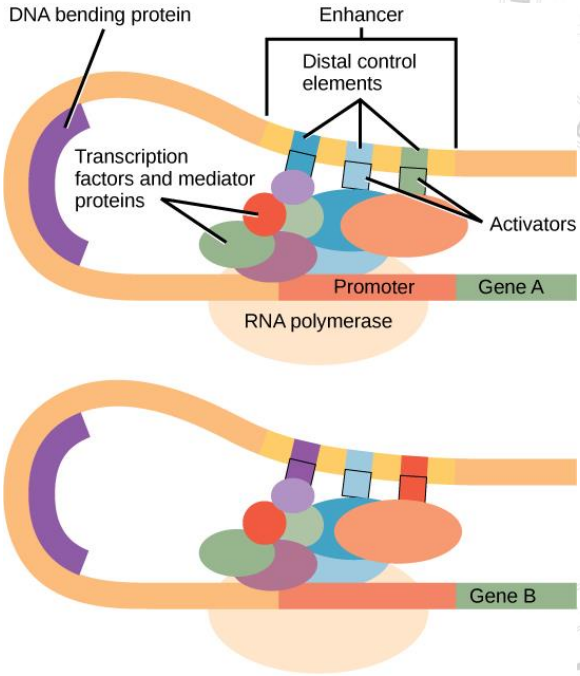


Figure 14. An enhancer is a DNA sequence that promotes transcription. It is recognized by activator proteins, which facilitate the recruitment and assembly of components of the pre-initiation complex, including general transcriptioin factors and RNA polymerase.

From https://courses.lumenlearning.com/wm-biology1/chapter/prokaryotic-transcription/

Eukaryotic cells also have mechanisms to prevent transcription. Transcriptional repressors can bind to enhancer regions and block transcription by preventing the binding of activating transcription factors and assembly of the pre-initiation complex.

Table 2 gives some examples transcriptional activators present in mammals, and the enhancer sequences at which they bind.

Transcription Activator	Consensus Sequence	Transcription Activator	Consensus Sequence
AP-1	TGAGTCA	**p53**	PuGPuCATGPyCPy
AP-2	CCC(A/C)N(C/G)$_3$	**NF-kB**	GGGPuNTPyPyCC
Oct-1	ATGCAAAT	**NFAT**	GGAGAPu
GATA-1	(A/T)GATAPu	**NF-E2**	TGACTCAG

*Pu = Purine (A or G); Py = Pyrimidine (C or T); N = any

The classic example of a transcriptional activator is p53. A tumor suppressor protein, it plays an important role in cell cycle control and apoptosis by regulating the expression of stress response and apoptosis genes such as *Bax* and *p21*. The p53 protein product is the most commonly mutated tumor suppressor in human cancers.

Epigenetics and gene expression in eukaryotes

Epigenetics is the study of inherited changes in phenotype or gene expression that does not involve a change in DNA sequence. It is an important mechanism by which environmental factors, such as diet, can influence biological processes and phenotypes, including disease.

Recall that eukaryotic DNA is tightly packaged into nucleosomes. Altering chromatin structure affects the accessibility of DNA to regulatory proteins. Epigenetic alterations that can affect nucleosomal structure include changes in DNA methylation and histone regulation

1. Histone modification

   Chromatin is held together by electrostatic interactions between positively charged lysines in histones and negatively charged nucleotides. Two enzymes regulate chromatin accessibility. Histone acetyltransferases (HATs) loosen histone – DNA interaction by removing lysine's positive charge, thereby enhancing transcription, while histone deacetylases (HDs or HDACs) repress transcription by returning lysine's negative charge.

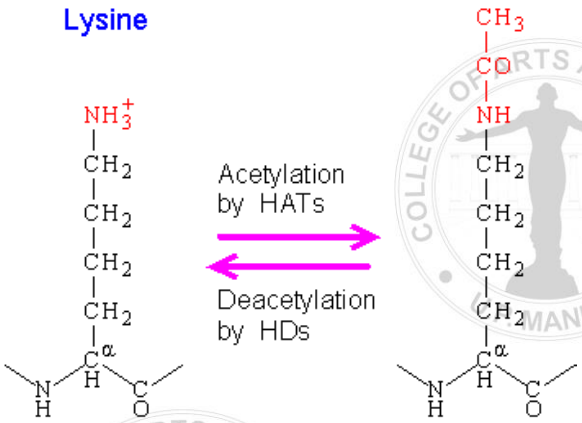


Figure 15. Acetylation and deacetylation of lysine residue. From https://www.web-books.com/MoBio/Free/Ch4G.htm

2. DNA methylation

   The presence of a methyl groups in cytosine does not change its basepairing activity. However, they present a steric hindrance to the binding of regulatory proteins when found in so – called CpG islands, which are DNA regions with a high frequency of CGs,

and are usually found in gene regulatory regions. Thus, methylated genes are usually silenced. Hypermethylation of promoter regions is associated with transcriptional silencing, and is a common mechanism for inactivation of tumor suppressor genes (TSGs) in human cancers. About half of TSGs mutated in familial cancer syndromes, such as RB1 (in retinoblastoma) and BRCA1 (in breast/ovarian cancer) have been shown to be inactivated by promoter methylation.

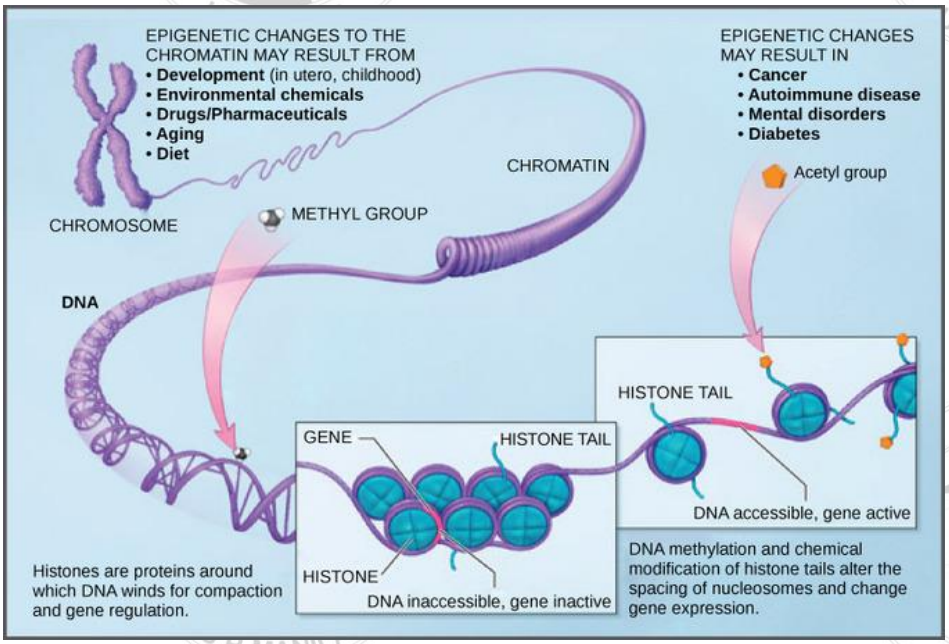


Figure 16: **Modifications to histones and DNA can alter gene expression.** Histone proteins and DNA nucleotides can be modified chemically. Modifications affect nucleosome spacing and gene expression. From https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_General_Biology_(Boundless)/16%3A_Gene_Expression/16.3%3A_Eukaryotic_Gene_Regulation/16.3C%3A_Epigenetic_Control%3A_Regulating_Access_to_Genes_within_the_Chromosome

Hypermethylated DNA regions with deacetylated histones are tightly coiled and transcriptionally inactive. These changes to DNA are inherited from parent to offspring, such that even if the DNA sequence itself is not altered, the pattern of transcriptional activation and gene expression is passed to the next generation. However, epigenetic changes may not necessarily be permanent. If the environmental pressure is removed, these may eventually fade. For example, fruit flies exposed to geldanamycin showed unusual outgrowths on their eyes that lasted through at least 13 generations of offspring, but were no longer observed in subsequent generations.

REFERENCES

Stryer 8th ed

https://courses.lumenlearning.com/boundless-biology/

https://www.nature.com/scitable/topicpage/transcription-factors-and-transcriptional-control-in-eukaryotic-1046/

https://cnx.org/contents/GFy_h8cu@11.10:dQV50wLv@7/Regulation-of-Gene-Expression

ADDITIONAL VIDEO RESOURCES

- Processing of Gene Information Prokaryotes versus Eukaryotes HD Animation
  https://www.youtube.com/watch?v=HBQrHZTq_IQ&list=PLYCGVJq0DVwKrmoSIvhzAOh0SpNUgfqFf&index=38
- Transcription and mRNA processing | Biomolecules | MCAT | Khan Academy
  https://www.youtube.com/watch?v=JQIwwJqF5D0
- Transcription Animation
  https://www.youtube.com/watch?v=vLz2A1cjPH8
- Eukaryotic transcription
  https://www.youtube.com/watch?v=JOBwqwxgJqc
- mRNA Synthesis Transcription HD Animation
  https://www.youtube.com/watch?v=lDOYSPkDaAM&list=PLYCGVJq0DVwKrmoSIvhzAOh0SpNUgfqFf&index=44
- Transcription Factors HD Animation
  https://www.youtube.com/watch?v=tUqz9amsdyI&list=PLYCGVJq0DVwKrmoSIvhzAOh0SpNUgfqFf&index=9
- Transcription Complex and Enhancers HD Animation
  https://www.youtube.com/watch?v=68REOTr7uE8&list=PLYCGVJq0DVwKrmoSIvhzAOh0SpNUgfqFf&index=10
- RNA Splicing HD Animation
  https://www.youtube.com/watch?v=4J8WgrPiMME&list=PLYCGVJq0DVwKrmoSIvhzAOh0SpNUgfqFf&index=27
- How Spliceosomes Process RNA HD Animation
  https://www.youtube.com/watch?v=twBylLWOSLU

- What is epigenetics? - Carlos Guerrero-Bosagna
  https://www.youtube.com/watch?v=_aAhcNjmvhc&t=12s
- Epigenetics
  https://www.youtube.com/watch?v=kp1bZEUgqVI