

Transcription and RNA Modification

By Dr. Hossein Noorollahi

October, 2024

Role of Transcription:

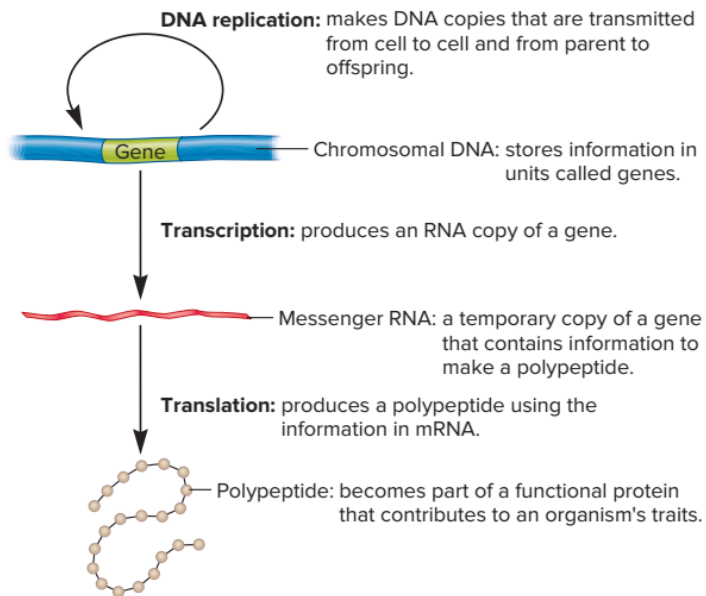
The **first step in gene expression**, crucial for protein synthesis and cellular function.

- The process of **transcription synthesizes RNA from a DNA template without altering the DNA structure**, allowing it to retain its information.
- Transcription produces messenger RNA (mRNA), which is then translated into polypeptides**, forming functional proteins that determine an organism's traits.

Central Dogma of Molecular Biology:

- Introduced by Francis Crick in 1958, it outlines the flow of genetic information: DNA → mRNA → Polypeptide.

Commented [hn1] درک مولکولی ما از رونویسی ژن‌ها به مطالعات مربوط به باکتری‌ها و باکتریوفاژها برمی‌گردد. در سال ۱۹۶۰ جاکوب و مزلسون دریافتند که پروتئین‌ها بر روی ریبوزوم‌ها سنتز می‌شوند. یک سال بعد جاکوب و مونود اظهار کردند که نوع خاصی از RNA به عنوان یک پیامبر ژنتیکی عمل می‌کند و اطلاعات لازم برای سنتز پروتئین را فراهم می‌کند. آنها فرض کردند این RNA که آن را mRNA نامیدند از روی توالی DNA رونویسی می‌شود و یک پلی‌پپتید خاص را سنتز می‌کند.



- **Replication vs. Transcription:**

- **Replication:**

Produces a complete copy of DNA

Replication enables genetic information to pass from one cell to another and from one generation to the next.

- **Transcription:**

Creates a copy of a specific gene

Transcription transfers information from the nucleus to the cytoplasm.

Key Components in Gene Expression (Transcription & Translation)

DNA Components: Essential for Transcription

- **Regulatory Sequences:**

- Binding sites for regulatory proteins.

- Influence transcription rates.
- Located in various regions of the DNA.
- **Promoter:**
 - Binding site for RNA polymerase.
 - Marks the **start of transcription.**
- **Terminator:**
 - Signals the end of transcription.

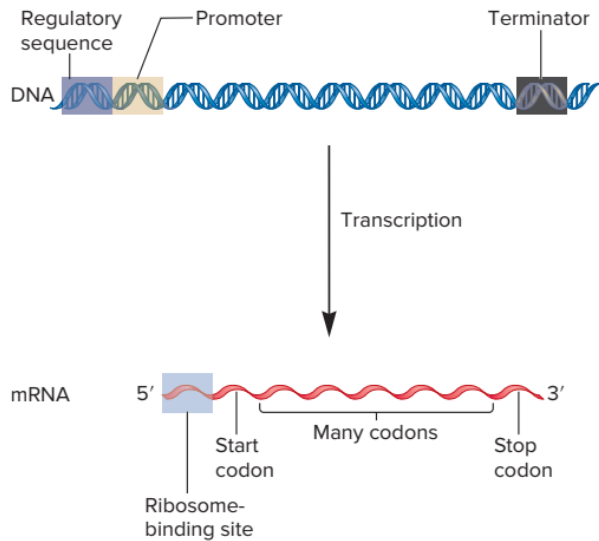
mRNA Components: Essential for Translation

- **Ribosome-Binding Site:**
 - Where ribosomes attach to mRNA.
 - Translation starts near this site; in eukaryotes, ribosomes scan for the start codon.
- **Start Codon:**
 - Indicates the first amino acid in a polypeptide.
 - formylmethionine in bacteria; methionine in eukaryotes.
- **Stop Codon:**
 - Signals the termination of polypeptide synthesis.

Commented [hn2]: توالی کوتاهی از DNA است که برای شروع رونویسی ضروری است. این توالی بازی موقعیت دقیق شروع رونویسی را مشخص می کند. بازهای موجود در توالی پروموتوری نسبت به جایگاه آغاز رونویسی شماره گذاری می شوند.

Commented [hn3]: جایگاه شروع رونویسی اولین بازی است که به عنوان الگو برای شروع رونویسی مورد استفاده قرار می گیرد و با مثبت ۱ علامت گذاری می شود. بازهای قبل از این جایگاه با علامت منفی شماره گذاری می شوند و هیچ بازی با صفر مشخص نمی شود. بنابراین بیشتر نواحی پروموتوری با شماره های منفی مشخص می شوند (بیشتر نواحی پروموتوری درست در بالا دست جایگاه آغاز رونویسی قرار دارند).

Commented [hn4]: وجود این دو توالی سبب می شود سنتز RNA در یک منطقه مشخص انجام شود. بنابراین فقط از حد فاصل بین پروموتر تا خاتمه دهنده رونویسی می شود



Abbreviations:

Template Strand:

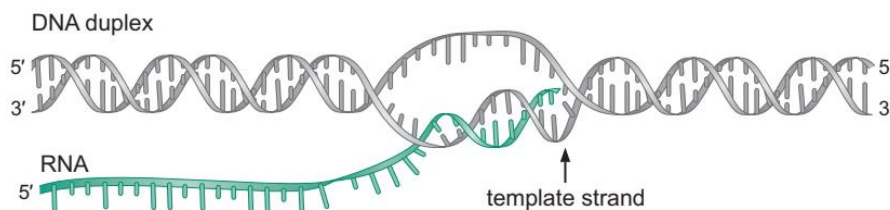
- The DNA strand used to synthesize mRNA.
- Complementary to the mRNA, differing by having thymine instead of uracil.

Commented [hn5] رشته‌ای که از روی آن ام آر آن ای ساخته می‌شود تمپلیت نامیده می‌شود که مکمل ام آر آن است با این تفاوت که به جای یو تی وجود دارد و رشته مقابل دینگ استرند نامیده می‌شود چون مشابه رونوشت است

Coding Strand:

- The opposite DNA strand that resembles the mRNA transcript.

Commented [hn6] برای ژن‌های کد کننده پروتئین، رشته غیر الگو را رشته رمزگذار یا سنس معادل Codind strand می‌نامند زیرا توالی آن مشابه توالی mRNA می‌باشد. همچنین، رشته الگو، رشته غیر کد کننده یا آنتی سنس نامیده می‌شود.



Codons:

- Three-nucleotide sequences that specify amino acids.
- The sequence of codons determines the amino acid sequence in polypeptides.

Polycistronic mRNA:

- In bacteria, can encode multiple polypeptides.

Directionality:

- **RNA polymerase synthesizes RNA in the 5' to 3' direction, similar to** how **DNA polymerase** replicates DNA.
- Nucleotides are added by **forming bonds between** the **5' phosphate group of a new nucleotide and the 3' hydroxyl group of the previous nucleotide.**

Complementarity Rule:

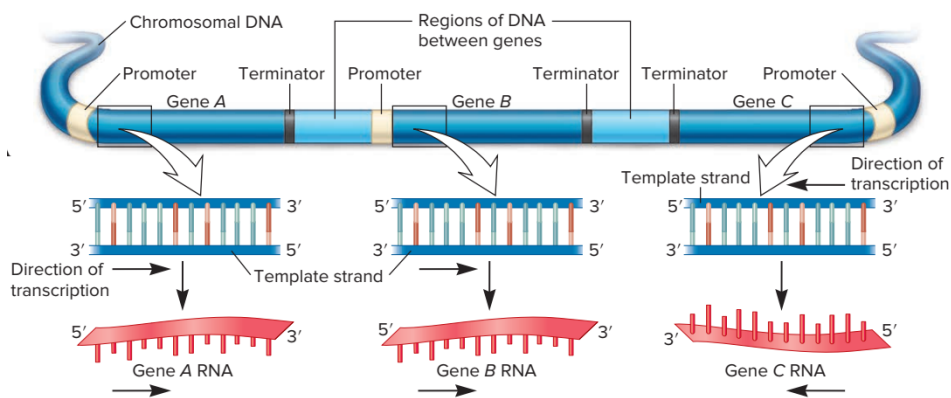
- The **base pairing rules during RNA synthesis are similar to** those in **DNA replication**, with **uracil (U) substituting for thymine (T)**:
 - A (adenine) pairs with U (uracil)
 - T (thymine) pairs with A (adenine)
 - G (guanine) pairs with C (cytosine)
 - C (cytosine) pairs with G (guanine)

Transcription of Multiple Genes

Direction of Transcription:

- The direction of transcription can vary among adjacent genes on a chromosome:
 - **Genes A and B:** Transcribed from left to right using the bottom DNA strand as the template.
 - **Gene C:** Transcribed from right to left using the top DNA strand as the template.

Commented [hn7] mRNA می تواند به صورت پلی سیسترون یا مونو سیسترون باشد. mRNA باکتری ها به صورت پلی سیسترونی بوده یعنی دو یا تعداد بیشتری پلی پپتید را کد می کند در حالی که mRNA یوکاریوتی به صورت مونوسیسترون بوده و از روی هر ژن فقط یک رونوشت ساخته می شود.



The Role of Promoters in Transcription Initiation

Promoter Definition: A **promoter** is a short DNA sequence crucial for initiating transcription. It "promotes" gene expression by **directing the exact location for transcription to begin**.

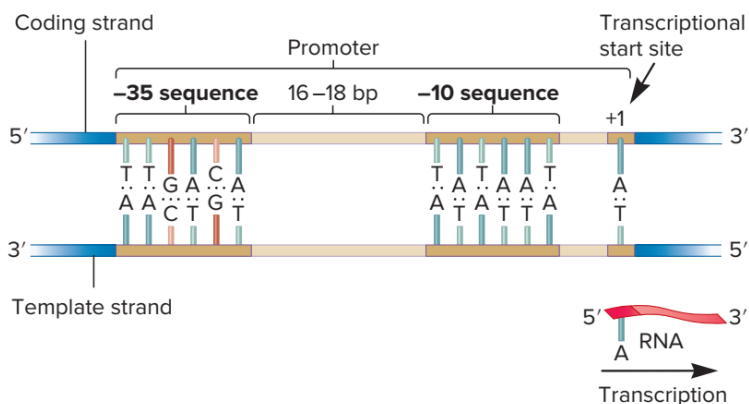
Location and Structure:

- The **promoter** is typically **located upstream of the transcription start site**, denoted as +1.
- Bases preceding **this site are numbered negatively** (e.g., -1, -2, -3, etc.).

Key Sequences in Promoters:

- In many *E. coli* promoters, two important sequences are found at approximately **the -35 and -10 positions**:
 - 35 Sequence:** 5'-TTGACA-3'
 - 10 Sequence (Pribnow Box):** 5'-TATAAT-3'
- The **Pribnow box** was discovered by **David Pribnow** in 1975 and is essential for transcription initiation.

- **Variations in the -35 and -10 sites among different genes can affect transcription efficiency**, with stronger promoters having sequences that closely match the consensus.



Commented [hn8] در بسیاری از ژن‌های باکتریایی همبستگی قوی بین حداکثر میزان رونویسی RNA با نواحی منهای ۱۰ و منهای ۳۵ موجود در توالی‌های حفاظت شده وجود دارد

Role of Proteins in Transcription:

- **Transcription factors** are proteins that regulate transcription rates by **binding to promoters and regulatory elements**.
- They can enhance or inhibit transcription, **controlling gene expression**.

RNA Polymerase Holoenzyme

RNA Polymerase Structure:

- **In E. coli, the core RNA polymerase enzyme consists of five subunits:** $\alpha 2$, β , β' , and ω

Commented [hn9] عوامل رونویسی یا TF توالی‌های بازی مشخصی را در DNA تشخیص می‌دهند و رونویسی را تنظیم می‌کنند (TF ها دارای دومین‌های متصل شونده به DNA هستند). برخی از این عوامل رونویسی مستقیماً به پروموتور متصل شده و رونویسی را تسهیل می‌کنند اما برخی دیگر توالی‌های تنظیمی مثل توالی‌های کوتاهی از DNA که در تنظیم بیان ژن نقش دارند را شناسایی می‌کنند. اتصال عوامل رونویسی به این توالی‌های تنظیمی می‌تواند میزان رونویسی را افزایش دهند و یا در برخی موارد رونویسی را مهار کنند.

Commented [hn10] آنزیمی که سنتز RNA را کاتالیز می‌کند RNA پلیمراز نامیده می‌شود.

Commented [hn11] در پروکاریوت ها مانند ای کولای، هسته آنزیم یا Core enzyme از ۵ زیر واحد شامل ۲ زیر واحد آلفا (36 kDa)، ۱ زیر واحد بتا (150 kDa)، ۱ زیر واحد بتا پرایم (155 kDa) و ۱ زیر واحد امگا تشکیل شده است.

TABLE 13-1 The Subunits of RNA Polymerases

Prokaryotic		Eukaryotic		
Bacterial Core	Archaeal Core	RNAP I (Pol I)	RNAP II (Pol II)	RNAP III (Pol III)
β'	A'/A''	RPA1	RPB1	RPC1
β	B	RPA2	RPB2	RPC2
α^I	D	RPC5	RPB3	RPC5
α^{II}	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	[+six others]	[+nine others]	[+seven others]	[+11 others]

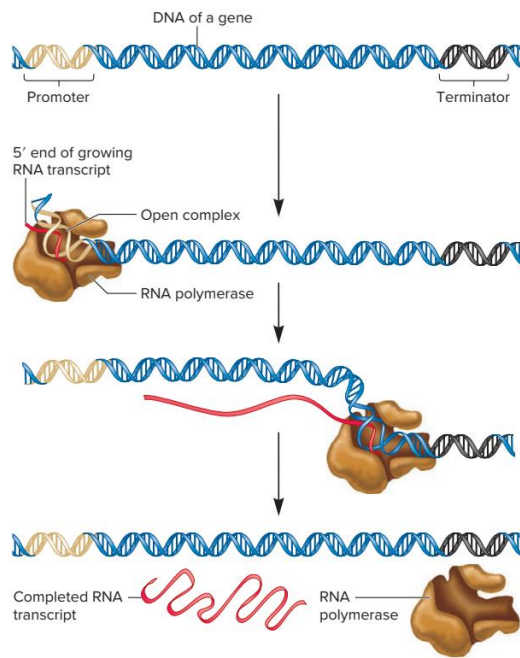
- The sigma (σ) factor associates with the core enzyme to form the RNA polymerase holoenzyme, which is essential for transcription initiation

Commented [hn12]: تجمع زیر واحد ششم به نام فاکتور سیگما با هسته آنزیم منجر به تشکیل هولوآنزیم می شود. تشکیل هولوآنزیم برای شروع رونویسی ضروری است و رونویسی با اتصال هولوآنزیم RNA پلیمراز به توالی پرموتر آغاز می شود.

Functional Roles of Subunits:

- α Subunits: Important for holoenzyme assembly and DNA binding
- β and β' Subunits: Catalyze RNA synthesis and bind to DNA
- ω Subunit: Aids in the assembly of the core enzyme
- Sigma (σ) factor: Plays a crucial role in recognizing the promoter sequences (-35 and -10)

Three Stages of Transcription: Initiation, Elongation, and Termination



Initiation: The promoter functions as a recognition site for transcription factors (not shown). The transcription factors enable RNA polymerase to bind to the promoter. Following binding, the DNA is denatured into a bubble known as the open complex.

Elongation/synthesis of the RNA transcript: RNA polymerase slides along the DNA in an open complex to synthesize RNA.

Termination: A terminator is reached that causes RNA polymerase and the RNA transcript to dissociate from the DNA.

1. Initiation:

Recognition Step:

- Transcription **begins when transcription factors recognize and bind to the promoter** region of the DNA.

Formation of the Closed Complex:

- The **holoenzyme binds loosely to the DNA**, sliding along until it encounters the promoter.
- The **σ factor recognizes the -35 and -10 sequences**, leading to the formation of a closed complex.

Formation of the Open Complex:

Commented [hn13]: توالی بازهای موجود در ناحیه پروموتور

توسط چندین عامل رونویسی شناسایی می‌شوند (اتصال اختصاصی عوامل رونویسی به توالی پروموتور جایگاه آغاز رونویسی را معین می‌کند).

Commented [hn14]: فرایند رونویسی زمانی آغاز می‌شود

که فاکتور سیگما موجود در هولوآنزیم به ناحیه پروموتوری متصل شده و تشکیل کمپلکس بسته Closed complex دهد.

Commented [hn15]: در ادامه دو رشته مارپیچ DNA باید باز

شده تا یکی از رشته‌ها بتواند به عنوان الگو برای سنتز رشته مکمل بکار رود که منجر به تشکیل ساختار حباب مانند کوچکی می‌شود که کمپلکس باز نامیده می‌شود. باز شدن DNA ابتدا در توالی TATAAT موجود در ناحیه منهای ۱۰ که دربردارنده جفت بازهای AT است انجام می‌شود. چون جدا کردن ناحیه غنی از جفت بازهای AT که دارای دو پیوند هیدروژنی هستند راحت‌تر است.

- **RNA polymerase, along with transcription factors, binds to the promoter while the DNA is still in its double-helix form.**
- **The DNA strands are then separated, creating a bubble-like structure known as the open promoter complex.**
- For transcription to proceed, **the DNA must unwind, starting at the TATAAT sequence in the -10 region, which is AT-rich and easier to separate due to fewer hydrogen bonds.**

بطور خلاصه مرحله شروع رونویسی را می توان به این صورت بیان کرد که زیر واحد فاکتور سیگما در هولوآنزیم RNA پلیمراز نواحی منهای ۱۰ و منهای ۳۵ پروموتور را تشخیص می دهد. DNA در ناحیه منهای ۱۰ باز شده و کمپلکس باز ایجاد می کند و به دنبال آن یک رشته کوتاه RNA سنتز می شود. سپس فاکتور سیگما از هولوآنزیم جدا شده و هسته آنزیمی RNA پلیمراز می تواند با حرکت روی DNA، RNA سنتز کند (تشکیل کمپلکس باز در این مرحله ادامه می یابد)

2. Elongation:

After the initiation of transcription, the **elongation stage involves the synthesis of the RNA transcript** as RNA polymerase progresses along the DNA template strand.

Once the **RNA polymerase** has **synthesized a short stretch of RNA** (≈ 10 bases), it **shifts into the elongation phase.**

- **Unwinds the DNA in front**
- **reanneals in behind**
- **growing RNA chain from the template**
- **performs proofreading functions**

RNA Synthesis:

- **Once the open complex is formed, a short RNA strand is synthesized, and the σ factor is released.**

Commented [hn16]: طویل سازی رشته RNA با جدا شدن

فاکتور سیگما دنبال می شود و هسته آنزیمی روی DNA می لغزد و سنتز رشته RNA را انجام می دهد

- **This release marks the transition from initiation to the elongation phase** of transcription, **allowing the core enzyme to synthesize RNA as it moves along the DNA.**
- **RNA polymerase moves along the DNA template strand, synthesizing a complementary RNA strand**
- **This process continues as RNA polymerase slides along the DNA,** elongating the RNA transcript.

3. Termination:

- Transcription concludes **when RNA polymerase reaches a terminator sequence in the DNA.**
- **Termination of transcription is a crucial step that marks the end of RNA synthesis.** This process involves the release of RNA polymerase and the newly synthesized RNA transcript from the DNA template.

Commented [hn17]: خاتمه زمانی رخ می دهد که هیبرید

کوتاه DNA - RNA از هم جدا شده که منجر به رها شدن RNA

پلیمراز و رونوشت RNA تازه سنتز شده می شود

Mechanisms of Termination

In *E. coli*, there are **two primary mechanisms** for transcription termination:

1. **ρ -Dependent Termination**
2. **ρ -Independent Termination**

1. ρ -Dependent Termination

Key Components:

- **Rho (ρ) Protein:** An **RNA-binding protein** that plays a central role in terminating transcription.
- **Rut Site:** A **specific sequence in the DNA** that encodes a corresponding sequence in the RNA, **acting as a recognition site for the binding of the ρ protein.**

Mechanism:

- **After the rut site is synthesized in the RNA, the ρ protein binds to the RNA and moves toward the RNA polymerase.**
- The **termination site contains a sequence in the DNA that encodes for an RNA sequence rich in GC base pairs, forming a stem-loop structure.**
- **This stem-loop forms shortly after its corresponding RNA sequence is synthesized and quickly binds to RNA polymerase.**
- **The binding induces a conformational change in RNA polymerase, causing it to pause.**
- **The ρ protein catches up to the paused RNA polymerase and disrupts the hydrogen bonds between the RNA and DNA within the open complex.**
- **This results in the release of the completed RNA strand and RNA polymerase from the DNA.**

Commented [hn18]: در این صورت پروتئین Rho به عنوان هلیکاز عمل کرده، نواحی هیبرید DNA - RNA را از هم جدا می کند

Commented [hn19]: سنتز RNA چندین نوکلئوتید فراتر از این ساختار خاتمه می یابد. چون ساختار ساقه حلقه باعث متوقف شدن RNA پلیمراز می شود.

2. ρ -Independent Termination

Key Features:

- This mechanism **does not require** the involvement **of the ρ protein.**
- It typically **involves two adjacent sequences in the DNA:**
 - **One sequence promotes the formation of a stem-loop structure.**
 - The **second sequence**, located downstream, is **rich in uracil (U)** at the **3' end of the RNA.**

Mechanism:

- The **formation of the stem-loop structure causes RNA polymerase to pause** in its synthesis.
- **Proteins such as NusA can bind to RNA polymerase, stabilizing this pause** at the stem-loop.
- When RNA polymerase pauses, **the uracil-rich sequence in the RNA transcript binds to the DNA template strand.**

- The **hydrogen bonds between the uracil-rich RNA and the DNA are relatively weak, leading to a spontaneous dissociation of the RNA transcript from the DNA.**
- This cessation of transcription **occurs without the need for the ρ protein**, hence the term **intrinsic termination**.

Approximately **half of E. coli genes utilize intrinsic termination, while the other half rely on the ρ protein for termination.**

Gene Transcription in Eukaryotes vs. Bacteria

1. Complexity in Eukaryotes:

- Eukaryotic transcription is **more complex due to larger cell size and the presence of organelles.**
- **Basic features** of gene transcription **are similar in both bacterial and eukaryotic species.**
- **Key insights are derived from studies in *Saccharomyces cerevisiae*** and other eukaryotes, including mammals.

Commented [hn20]: رونویسی در یوکاریوت‌ها نیازمند تغییر در موقعیت و ساختار نوکلئوزوم است

2. RNA Polymerases in Eukaryotes:

- Eukaryotic cells utilize three RNA polymerases:
 - **RNA Polymerase I: Transcribes ribosomal RNA (rRNA), excluding 5S rRNA.**
 - **RNA Polymerase II: Transcribes all protein-coding genes and most small nuclear RNAs (snRNAs).**
 - **RNA Polymerase III: Transcribes tRNA genes and the 5S rRNA gene, with some involvement in non-coding RNAs.**
- Magnesium ions (Mg^{2+}) play a crucial role at the catalytic site, enabling the attachment of nucleotides to the growing RNA strand

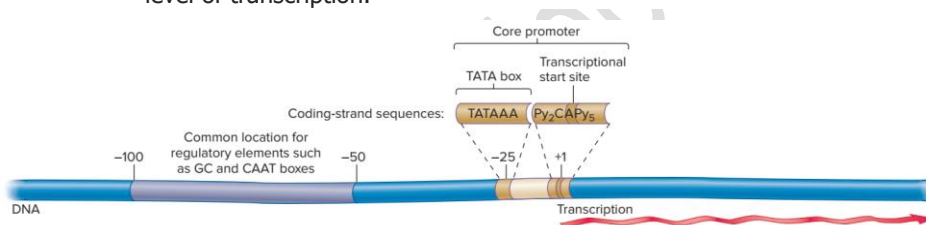
Eukaryotic Gene Regulation

Key Features of Eukaryotic Genes:

- **Eukaryotic genes consist of** a **core promoter** and **regulatory elements**, essential for proper transcription rates.

Core Promoter:

- The **core promoter is a short DNA sequence crucial for initiating transcription**.
- It typically **includes the TATA box (TATAAA), located about 25 base pairs upstream of the transcription start site**.
- The **TATA box defines the precise start point for transcription**; its absence leads to undefined start sites and variable transcription initiation.
- **The core promoter alone results in basal transcription**, which is a low level of transcription.



Regulatory Elements:

- Regulatory elements are **short DNA sequences** that **modulate RNA polymerase's ability to recognize the core promoter**.
- Commonly **found in the -50 to -100 region**, but their positions can vary significantly across different genes.

Types of Regulatory Elements:

- **Enhancers:** Activating sequences that stimulate transcription. **Without enhancers, eukaryotic genes exhibit very low basal transcription levels.**
- **Silencers:** **DNA sequences that inhibit transcription** when bound by specific transcription factors.

Commented [hn21]: اگر این ناحیه از پروموتور حذف شود جایگاه شروع رونویسی نامشخص می ماند و احتمالاً رونویسی از جایگاه های متنوع آغاز می شود

These elements interact with transcription factors, which are proteins that influence transcription rates.

Cis-acting vs. Trans-acting Elements:

- **Cis-acting elements** (e.g., TATA box, enhancers, silencers) are **located on the same chromosome as the genes they regulate**.
- **Trans-acting factors** are **regulatory proteins encoded by separate genes**, which **may be on different chromosomes**.

Initiation of Transcription in Eukaryotic Protein-Encoding Genes

Proteins Involved in Transcription (Eukaryotic)

- RNA polymerase II
- General transcription factors (GTFs)
- Mediator complex

1. RNA Polymerase II:

- The enzyme responsible for **catalyzing** the linkage of **nucleotides in the 5' to 3' direction using DNA as a template**.
- **Composed of 12 subunits**, with the two largest subunits resembling the β and β' subunits of E. coli RNA polymerase.

2. General Transcription Factors (GTFs):

- **Critical for the initiation of transcription** at the core promoter.
- **TFIID:**
 - Recognizes and Binds to the TATA box**, crucial for recognizing the core promoter
 - Composed of** TATA-binding protein (**TBP**) and TBP-associated factors (**TAFs**)
- **TFIIB:**

Binds to TFIID and facilitates the binding of RNA polymerase II to the core promoter.

Promotes the binding of TFIIF.

- **TFIIF**: Assists in **stabilizing RNA polymerase II**
Facilitates the **binding of TFIIE and TFIH to RNA polymerase**
- **TFIIE and TFIH**: **Complete the assembly of** proteins into a **preinitiation complex**

TFIH:

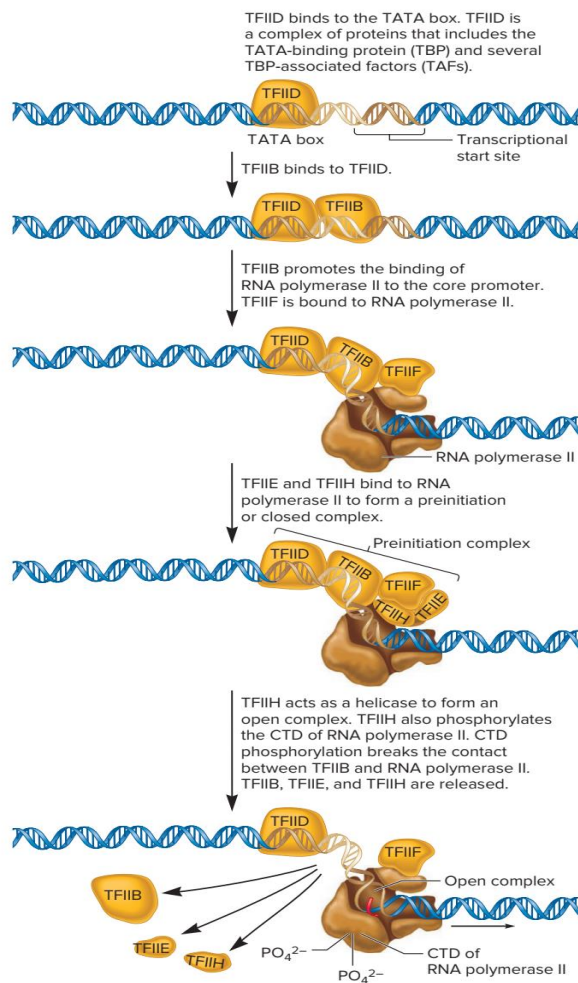
- **Multisubunit protein with multiple functions:**
Acts as a helicase to promote the formation of the **open complex** **(Contains helicase subunits that unwind DNA strands)**
Plays a critical role in converting the closed complex into an open complex

Phosphorylates the carboxyl terminal domain (CTD) of RNA polymerase II (Hydrolyzes ATP), releasing its interaction with TFIIB, allowing progression to elongation.

Phosphorylation → releases TFIIB from RNA polymerase II, allowing it to proceed to elongation

TFIIE:

- **Involved in forming and maintaining the open complex** during transcription initiation.
- **Helps** regulate the binding and **activity of TFIH.**



2. Basal Transcription Apparatus:

- TFIID, TFIIB, TFIIF, TFIIE, TFIIH
- RNA polymerase II
- With a TATA box-containing DNA

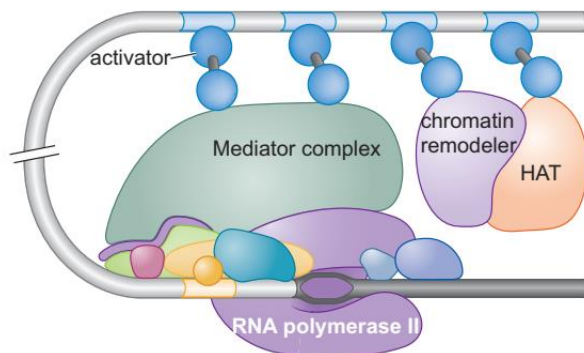
These components together form the **basal transcription apparatus**

3. Role of Mediator:

Commented [hn22]: توسط کورنبرگ و همکاران در ۱۹۹۰

کشف شد

- The **mediator complex**, discovered by Roger Kornberg, A **multisubunit complex** that **mediates interactions between RNA polymerase II and regulatory transcription factors** bound to enhancers or silencers
- It serves as an interface for diverse regulatory signals and **is crucial for the transition from transcription initiation to elongation**
- Mediator **can also phosphorylate the CTD of RNA polymerase II and regulate TFIIF's phosphorylation activity**



Transcriptional Termination in Eukaryotic Genes

1. Polyadenylation:

- **Eukaryotic** mRNAs undergo a **modification process** known as **polyadenylation**, which **involves cleavage near the 3' end followed by the addition of adenine nucleotides**.

2. **Polyadenylation Signal (PAS)** is a **key sequence in eukaryotic genes**, with the most common sequence being AAUAAA, active in mRNA processing

Definition: The **polyA signal sequence** is a **specific nucleotide sequence that signals the end of transcription** and the addition of a poly(A) tail to the mRNA molecule.

Location: This signal is typically found **near the 3' end of the gene**, often **represented by the sequence AAUAAA** or similar variants.

 **Function and Importance** This signal is **essential for mRNA stability, protecting it from degradation and facilitating nuclear export**

- **Termination of Transcription:** When RNA polymerase transcribes the DNA, it recognizes the polyA signal, which leads to the cleavage of the newly synthesized pre-mRNA.
- **Polyadenylation:** After cleavage, a poly(A) tail (a stretch of adenine nucleotides) is added to the 3' end of the mRNA. This modification is critical for:
 - **Stability:** The poly(A) tail protects the mRNA from degradation.
 - **Export:** It facilitates the export of the mRNA from the nucleus to the cytoplasm.
 - **Translation:** It enhances the efficiency of translation by aiding in the recognition of the mRNA by ribosomes.
- The **polyadenylation signal** sequence **directs the cleavage of the mRNA**, typically **occurring 500–2000 nucleotides downstream from this signal**.
- **This cleavage takes place before the termination of transcription.**

3. Models of Transcriptional Termination:

- **Two primary models** have been proposed to explain how transcriptional termination occurs:

a. Allosteric Model:

- **RNA polymerase II destabilized after transcribing the polyA signal sequence.**
 - **Destabilization may result from:**
 - The **release of elongation factor proteins**
 - The **binding of termination factor proteins**

b. Torpedo Model:

- Suggests that **RNA polymerase II is physically removed from the DNA.**

- In this model, the **RNA in downstream of the polyadenylation signal is cleaved by an exonuclease, which degrades the transcript in the 5' to 3' direction.**
- **When the exonuclease catches up to RNA polymerase II, it triggers the dissociation of RNA polymerase II from the DNA.**

RNA Modifications in Eukaryotes

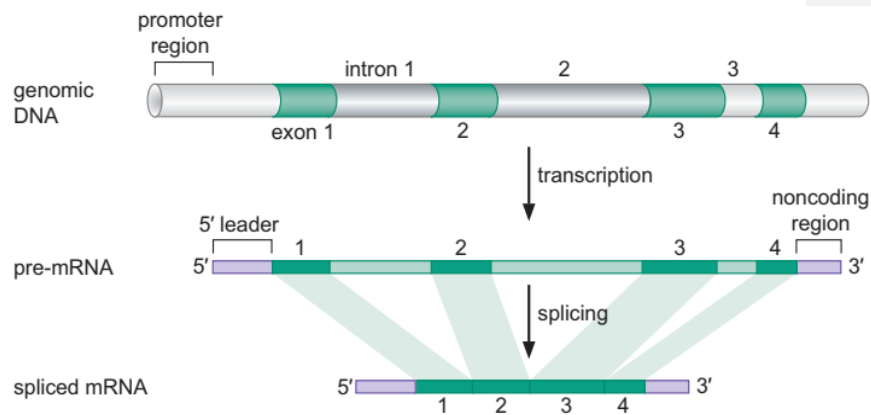
Includes:

1. Removing of Introns,
2. Adding 5' Capping
3. Adding 3' PolyA Tail

- In the late 1970s, advancements in molecular biology revealed that **eukaryotic protein-encoding genes are not always colinear with their functional mRNAs.**
- **Eukaryotic genes contain coding sequences** (exons) interrupted by **non-coding sequences** (introns).
- **Exons:** Coding regions that are retained in the mature mRNA.
- **Introns:** Non-coding regions that are **removed during RNA processing.**

Primary Transcript:

- **Eukaryotic genes are transcribed into a primary RNA** transcript (pre-mRNA) that **includes both introns and exons.**
- Most **protein-encoding genes** in complex eukaryotes **contain one or more introns**, while introns can also be found in tRNA and rRNA genes.

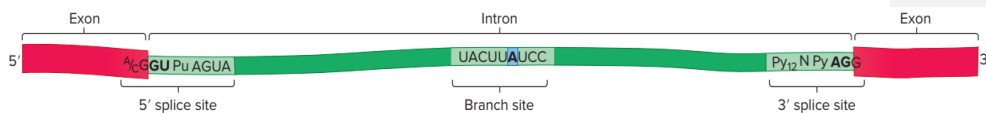


RNA Splicing Mechanisms

- **RNA splicing** is crucial for **converting pre-mRNA into mature mRNA** by precisely **removing introns and joining exons**.
Includes: Pre-mRNA Splicing & Self-Splicing Introns

Important Sequences:

1. **5' Splice Site:** Marks the **start of an intron**, recognized by the **conserved sequence "GU."**
2. **3' Splice Site:** Marks the **end of an intron**, recognized by the **conserved sequence "AG."**
3. **Branch Site:** Located within the intron, typically **near the 3' end**, identified by a **conserved "A"** and followed by a polypyrimidine tract (Py tract).



The GU at the 5' splice site, AG at the 3' splice site, and A at the branch site are highly conserved

The Role of the Spliceosome in pre-mRNA Splicing

- Unlike self-splicing introns, pre-mRNA splicing requires a complex known as the **spliceosome**.
- The **spliceosome recognizes intron-exon boundaries and facilitates the removal of introns, ensuring proper mRNA maturation before it exits the nucleus.**
- RNA splicing is facilitated by the spliceosome, a large molecular complex composed of **proteins** and **small nuclear RNAs** (snRNAs).

Key Components:

- **Spliceosome Structure:**
 - **Comprises** approximately **150 proteins and five snRNAs** (U1, U2, U4, U5, U6).
 - **Similar** in size **to ribosomes**, which synthesize proteins.
 - **Hydrolyzes multiple ATP** molecules during splicing.
- **Function of snRNAs:**
 - Each **snRNA** is **100-300 nucleotides** long and **forms complexes with proteins**, known as **small nuclear ribonucleoproteins** (**snRNPs**).
 - **Functions** primarily include **recognizing splice sites, bringing them together, and catalyzing RNA cleavage and joining.**

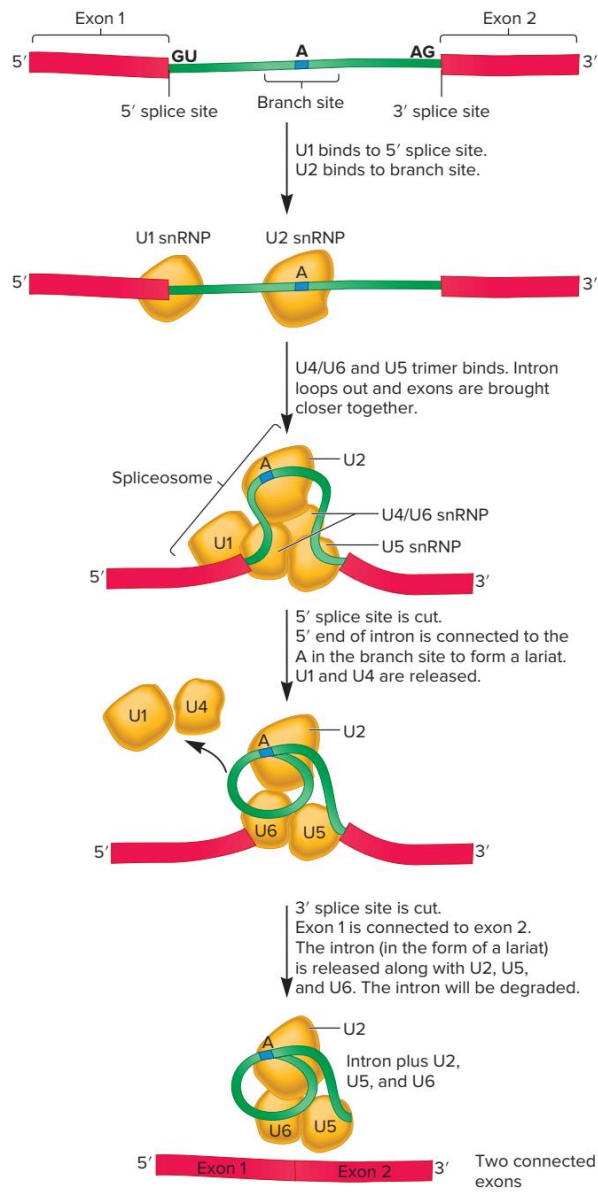
Splicing Mechanism:

1. **Recognition of Splice Sites:**
 - **U1 snRNA** binds to the **5' splice site** via complementary base pairing.
 - **U2 snRNA** recognizes the **branch site**.
 - **U6 snRNA** interacts with the **5' splice site later** in the process.
2. **Bridging Interactions:**
 - **U2 and U6** snRNAs interact to **bring the 5' splice site and branch site together**, facilitating the splicing reaction.

- The remaining snRNPs (**U2, U5, and U6**) remain **attached to** the **intron**, which **forms** a **lariat structure before** being **degraded**
- A **trimer of snRNPs (U4/U6 dimer + U5)** binds next, leading to the looping out of the intron and **bringing the exons closer together**

3. Role of **Non-snRNP Proteins**:

- **U2AF**: **Recognizes** the **polypyrimidine** tract at the **3' splice site** and **assists** in the initial **binding** of **other proteins**
- **Branchpoint-Binding Protein (BBP)**: **Binds** to the **branch site before** being replaced by **U2 snRNP**.
- **DEAD-box Helicases**: **Facilitate RNA rearrangements** and help in the **removal of spliced mRNA** from the spliceosome.

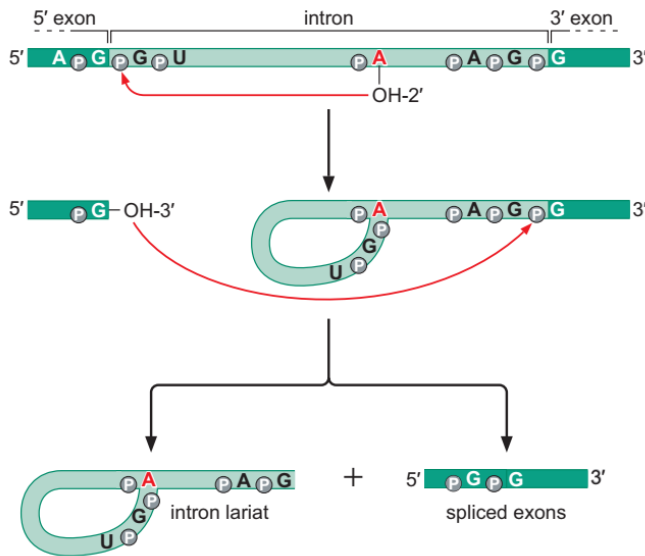


Lariat Formation in RNA Splicing

- The **splicing** process **involves** the **removal of introns and the joining of** flanking **exons through a series of chemical reactions.**

Key Steps in Splicing:

- **Transesterification Reactions: Two successive reactions break and form phosphodiester bond** within the pre-mRNA
 - Two bonds are broken, and two are formed
- **First Reaction:**
 - **Triggered by** the **2'-OH of the conserved A at the branch site.**
 - **Acts as a nucleophile, attacking the phosphoryl group of G at the 5' splice site.**
 - Result: **Cleavage of the 5' junction and formation of a three-way junction, creating lariat structure.**
- 2. **Second Reaction:**
 - The **newly liberated 3'-OH of 5' exon acts as a nucleophile, attacking the phosphoryl group at the 3' splice site.**
 - Result: **Joins the 5' and 3' exons together** and **releases the intron** in the lariat form.



Self-Splicing Introns

Self-splicing introns **demonstrate that RNA can catalyze its own splicing**, revealing the catalytic potential of RNA molecules.

- **Group I and Group II introns can remove themselves from RNA without external proteins.**

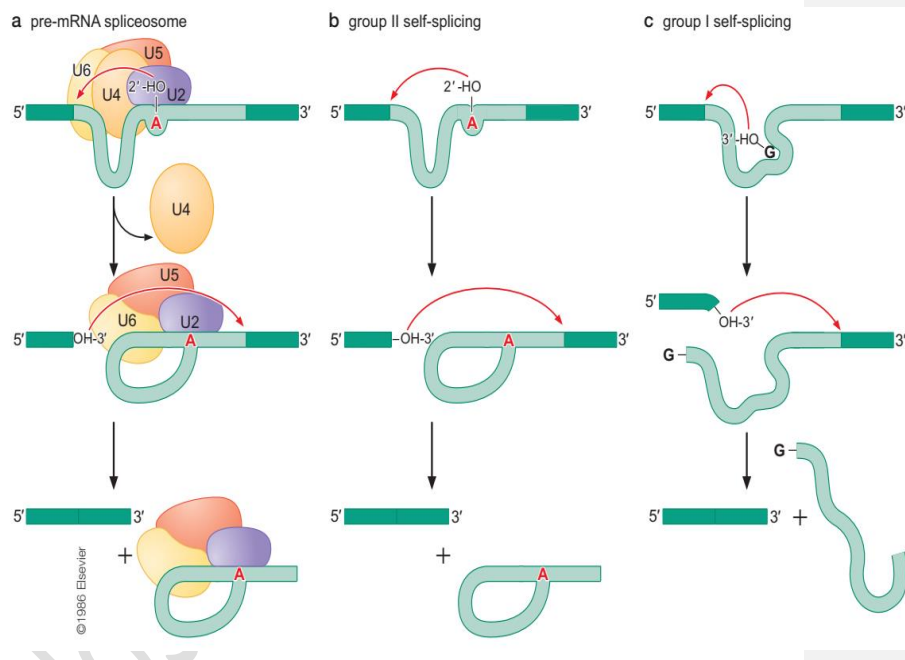
Group I Introns:

- **Mechanism:**
 - **Utilize a free guanine nucleotide in branch site** to initiate splicing.
 - The **guanine binds to a pocket in the RNA, allowing its 3'-OH group to attack the 5' splice site.**
 - This **results** in the formation of a **linear intron** and the **joining of exons.**

Group II Introns:

- **Mechanism:**

- **Similar to spliceosome-mediated splicing, using an adenine residue at the branch site.**
- **Forms a lariat** structure during the splicing process.



Alternative Splicing Regulates Exon Inclusion in Mature mRNA

- A mechanism that allows a single gene to produce multiple mRNA isoforms by varying the combination of exons included in the final mRNA.

1. Advantages of Alternative Splicing:

- **Allows a single gene to produce multiple polypeptides with different amino acid sequences,** potentially altering their functions.

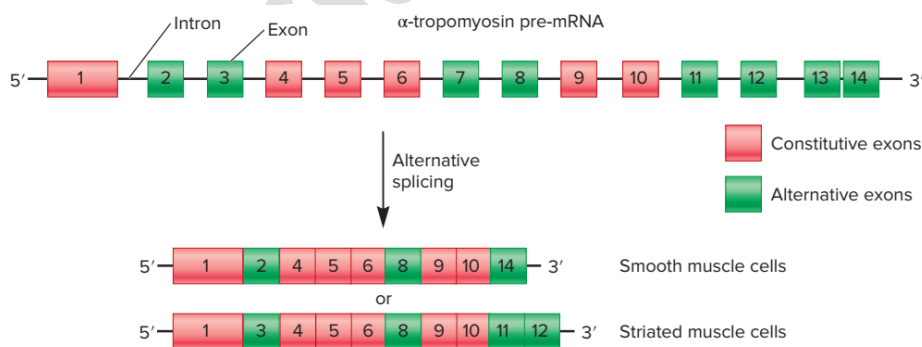
- **Most alternative protein versions retain similar functions due to shared amino acid sequences**, but unique characteristics arise from differences.
- **Enables organisms to maintain fewer genes while increasing protein diversity.**

2. Variation Among Species:

- **In unicellular organisms** like *Saccharomyces cerevisiae* (Baker's yeast), **only about 5% of genes are spliced**, with minimal alternative splicing.
- **In contrast, complex multicellular organisms**, such as humans, rely heavily on alternative splicing, **with about 70% of pre-mRNAs being alternatively spliced.**

3. Case Study: α -Tropomyosin:

- It **contains 14 exons, with 6 constitutive exons present in all cell types, while alternative exons vary based on cell type.**
- For example, **smooth muscle cells include exon 2, while striated muscle cells include exon 3**, tailoring the protein's function to the specific needs of each cell type.



4. Regulation of Alternative Splicing:

- **Alternative splicing is not random;** it is **regulated by splicing factors.**
- **SR proteins** are **key splicing factors that influence splice site selection** through their RNA-binding and protein-protein recognition capabilities.

5. Mechanisms of Splicing Regulation:

- **Splicing factors** can **act as repressors or enhancers:**
 - **Repressors inhibit spliceosome recognition of splice sites, leading** to events like **exon skipping** (e.g., skipping exon 2).
 - **Enhancers promote spliceosome recognition, ensuring inclusion of specific exons** (e.g., including exon 3).

Pre-mRNA Modifications, to be Continued

Eukaryotic pre-mRNAs also undergo significant **modifications at** both their **5' and 3' ends, enhancing their stability and functionality.**

1. 5' Capping:

- Most **mature mRNAs feature** a **7-methylguanosine cap at the 5' end,** a process **known as capping.**
- Capping **occurs during transcription by RNA polymerase II when the transcript is** approximately **20–25 nucleotides long.**
- **Capping Process:**
 - a. **Removal of a Phosphate:** RNA **5'-triphosphatase removes** one **phosphate from the 5' end.**
 - b. **Attachment of Guanosine:** **Guanylyltransferase attaches** a **guanosine monophosphate (GMP) to the 5' end** using GTP.
 - c. **Methylation:** **Methyltransferase adds a methyl group to the nitrogen at position 7 of the guanine base.**

Functions of the 5' Cap:

- The **cap is recognized by cap-binding proteins, which play several** critical **roles:**

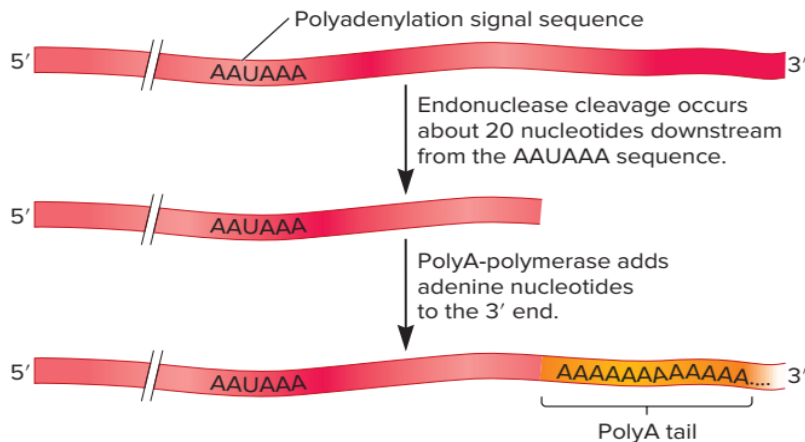
- a. **Nuclear Export:** **Required for** the proper **exit of mRNAs from the nucleus.**
- b. **Translation Initiation:** **Essential for** the early stages of **translation.**
- c. **Splicing Efficiency:** May **enhance the efficient splicing of introns,** particularly the first intron near the 5' end.

2. 3' PolyA Tail:

- The **3' end of** most **mature mRNAs contains** a **polyA tail, a string of adenine nucleotides.**
- This tail **is crucial for:**
 - a. **mRNA Stability:** **Protects mRNA from degradation** in the cytosol.
 - b. **Nuclear Export:** **Facilitates** the **exit of mRNA from the nucleus.**
 - c. **Translation:** **Enhances** the **synthesis of polypeptides.**

3. Polyadenylation Process:

- The **polyA tail** is not encoded in the gene sequence but **is added enzymatically after transcription.**
- **Steps of Polyadenylation:**
 - a. **Polyadenylation Signal:** The **pre-mRNA contains a signal sequence** (e.g., AAUAAA) **near its 3' end.**
 - b. **Cleavage:** **An endonuclease recognizes the signal and cleaves the pre-mRNA about 20 nucleotides downstream of the AAUAAA sequence.**
 - c. **Degradation:** The **RNA fragment beyond the cleavage point is degraded.**
 - d. **Tail Addition:** **PolyA-polymerase adds adenine nucleotides to form the polyA tail, typically around 250 nucleotides long.**



4. Conclusion:

- **The 5' cap and 3' polyA tail are vital modifications that enhance the stability, export, and translational efficiency of eukaryotic mRNAs, underscoring their importance in gene expression regulation.**

5. Comparison with Bacterial RNAs:

- **Some bacterial RNAs are also polyadenylated, but in bacteria, polyadenylation targets RNA for degradation in a structure called the **degradosome****

RNA Editing and Its Functional Significance

RNA editing is a critical post-transcriptional modification that enhances protein diversity and functionality.

1. Introduction to RNA Editing:

- **RNA editing** refers to the process of modifying the nucleotide sequence of an RNA molecule, which can **involve additions, deletions, or conversions of nucleotides.**
- This process can **affect mRNAs by generating start codons, stop codons, or altering the coding sequence of polypeptides.**

2. Discovery and Context:

- **RNA editing was first discovered in trypanosomes**, the protists responsible for sleeping sickness.
- Initially met with skepticism, subsequent **research confirmed that RNA editing occurs in various organisms and is functionally significant.**

3. Mechanisms of RNA Editing:

- **Uric Acid Addition/Deletion:** **In trypanosomes, editing often involves the addition or deletion of uracil nucleotides.**
- **Base Conversion:** A **more widespread mechanism** involves **deamination**, where **an amino group is removed from a base:**
 - **Cytosine to Uracil:** **When cytosine is deaminated, it converts to uracil.**
 - **Adenine to Hypoxanthine:** **When adenine is deaminated, it converts to hypoxanthine, which is recognized as guanine during translation.**

4. Example: Apolipoprotein B Editing:

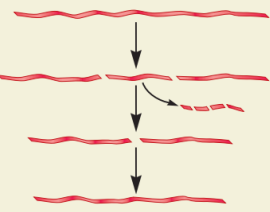
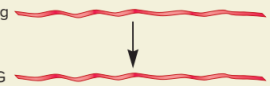
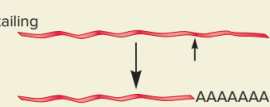
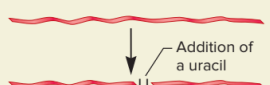
- **In mammals, RNA editing can significantly impact protein function.**
- **An example** involves the **mRNA for apolipoprotein B:**
 - **Unedited mRNA:** **In the liver, produces apolipoprotein B-100, crucial for cholesterol transport.**
 - **Edited mRNA:** **In intestinal cells, a single C is converted to a U, changing the codon from glutamine (CAA) to a stop codon (UAA).**
 - **Result:** **A shorter form of apolipoprotein B, demonstrating how editing can yield different proteins from the same gene.**

5. Comparison with Alternative Splicing:

- **RNA editing and alternative splicing both allow for the production of multiple protein variants from a single gene, increasing protein diversity.**

6. Prevalence of RNA Editing:

- In invertebrates like **Drosophila**, it is estimated that **50–100 pre-mRNAs undergo editing** that alters their coding sequences.
- In mammals, **fewer than 25 genes are currently known to have edited pre-mRNAs.**

Splicing		Splicing involves both cleavage and joining of RNA molecules. The RNA is cleaved at two sites, which allows an internal segment of RNA, known as an intron, to be removed. After the intron is removed, the two ends of the RNA molecules are joined together.	Splicing is common among eukaryotic pre-mRNAs, and it also occurs occasionally in rRNAs, tRNAs, and a few bacterial RNAs.
5' capping		The attachment of a 7-methylguanosine cap (m ⁷ G) to the 5' end of mRNA. The cap plays a role in the splicing of introns, the exit of mRNA from the nucleus, and the binding of mRNA to the ribosome.	Capping occurs on eukaryotic mRNAs.
3' polyA tailing		The attachment of a string of adenine-containing nucleotides to the 3' end of mRNA at a site where the mRNA is cleaved (see upward-pointing arrow). It is important for RNA stability and translation in eukaryotes.	Occurs on eukaryotic mRNAs and occasionally occurs on bacterial RNAs.
RNA editing		The change of the base sequence of an RNA after it has been transcribed.	Occurs occasionally in eukaryotic RNAs.

Key Differences Between Transcription and RNA Modification in Bacteria and Eukaryotes*

Component	Bacteria	Eukaryotes
Promoter	Consists of –35 and –10 sequences	For protein-encoding genes, the core promoter often consists of a TATA box and a transcriptional start site.
RNA polymerase	A single RNA polymerase	Three types of RNA polymerases; RNA polymerase II transcribes protein-encoding genes.
Initiation	σ factor is needed for promoter recognition.	Five general transcription factors assemble at the core promoter.
Elongation	Requires the release of σ factor	Mediator controls the switch to the elongation phase.
Termination	p-dependent or p-independent	According to the allosteric or torpedo model
Splicing	Very rare; self-splicing	Commonly occurs in protein-encoding pre-mRNAs in complex eukaryotes via a spliceosome; self-splicing occurs rarely.
Capping	Does not occur	Addition of 7-methylguanosine cap
Tailing	Added to 3' end; promotes degradation	Added to the 3' end; promotes stability
RNA editing	Does not occur	Occurs occasionally