

TRANSCRIPTION

Mechanism, Enzymology, and a Detailed Comparison of Prokaryotic and Eukaryotic Transcription

1. Introduction

Transcription is the process by which the genetic information encoded in a segment of DNA is copied into a complementary strand of RNA. It is the first step of gene expression, converting the cell's permanent, stable archive (DNA) into a mobile, disposable working copy (RNA) that can be read by ribosomes or that itself performs a structural or catalytic role.

- Carried out by **RNA polymerase** enzymes, which build an RNA strand complementary and antiparallel to a DNA template strand, following Watson–Crick base-pairing rules (with **U** incorporated opposite **A** instead of T).
- Unlike DNA polymerases, RNA polymerases can initiate a new chain *de novo* — they do not require a pre-existing primer.
- Only a short stretch of DNA is transcribed at a time, and only **one of the two DNA strands** (the template/antisense strand) is actually copied for any given gene; the other (coding/sense) strand has the same sequence as the resulting RNA (with T in place of U).
- Three main classes of RNA are produced by transcription: **messenger RNA (mRNA)**, which carries protein-coding information to ribosomes; **ribosomal RNA (rRNA)**, structural and catalytic component of ribosomes; and **transfer RNA (tRNA)**, which delivers amino acids during translation.

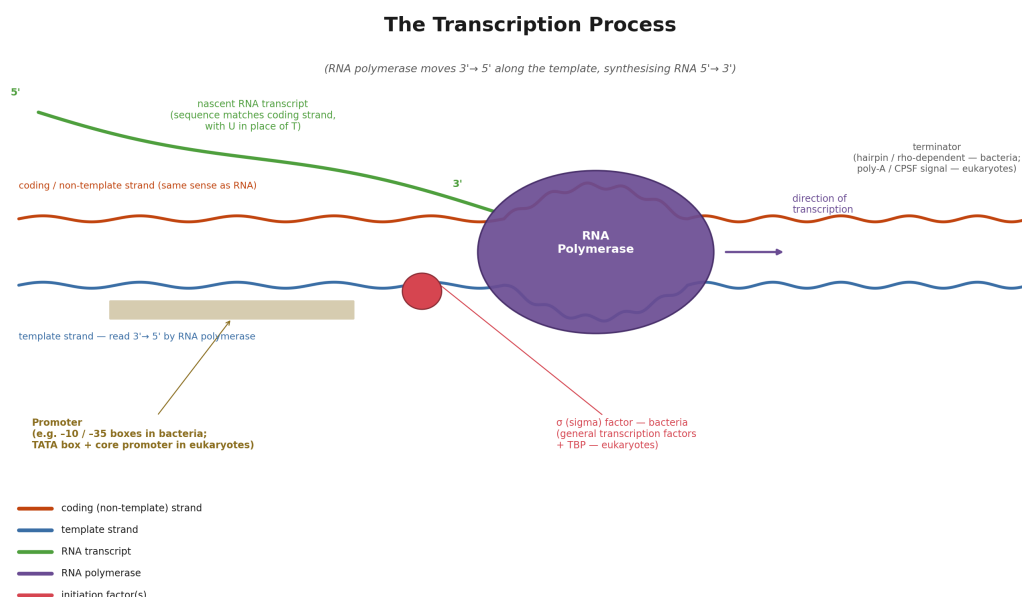


Fig. 1 — The transcription bubble: RNA polymerase unwinds the DNA duplex, reads the template strand 3'→5', and synthesises RNA 5'→3', guided to the promoter by initiation factors.

The Three Stages of Transcription

- Initiation** — RNA polymerase (assisted by additional factors) recognises and binds a specific DNA sequence called the **promoter**, which lies just upstream of the gene, and the DNA duplex is locally unwound to expose the template strand.

- **Elongation** — RNA polymerase moves along the template strand in the 3'→5' direction, synthesising a complementary RNA strand in the 5'→3' direction; the DNA duplex re-forms behind the moving enzyme as the small transcription bubble travels along the gene.
 - **Termination** — RNA polymerase reaches a specific sequence (a **terminator**) that signals it to stop, releasing both the completed RNA transcript and the DNA template.
-

2. Prokaryotic Transcription (e.g. *Escherichia coli*)

Bacteria use a single type of RNA polymerase to transcribe all classes of RNA. Because there is no nuclear envelope, transcription and translation are physically coupled — ribosomes can begin translating an mRNA while it is still being synthesised.

(a) RNA Polymerase Structure

- The bacterial **core enzyme** has the subunit composition $\alpha_2\beta\beta'\omega$: two α subunits (assembly and regulatory contacts), one β subunit and one β' subunit (forming the catalytic centre that binds DNA/RNA and catalyses bond formation), and one ω subunit (assists assembly and stability).
- The core enzyme alone can elongate an RNA chain but cannot specifically recognise a promoter.
- Addition of a **sigma (σ) factor** converts the core enzyme into the **holoenzyme**, which can locate and bind promoters correctly. Different sigma factors (e.g. σ^{70} , the housekeeping factor in *E. coli*) direct RNA polymerase to different sets of genes.

(b) Initiation

- Promoters in bacteria are typically described by two conserved short sequences upstream of the transcription start site (position +1): the **−35 box (consensus TTGACA)** and the **−10 box or Pribnow box (consensus TATAAT)**, spaced roughly 16–19 bp apart.
- The sigma factor recognises these elements and helps position the holoenzyme correctly, initially forming a **closed complex** (double-stranded DNA still intact) that isomerises into an **open complex** once the AT-rich −10 region is melted, exposing the template strand.
- After synthesising a short initial stretch of RNA (~8–9 nucleotides), the sigma factor is typically released, and the core enzyme proceeds into elongation alone.

Initiation Complex: Prokaryotic vs Eukaryotic RNA Polymerase

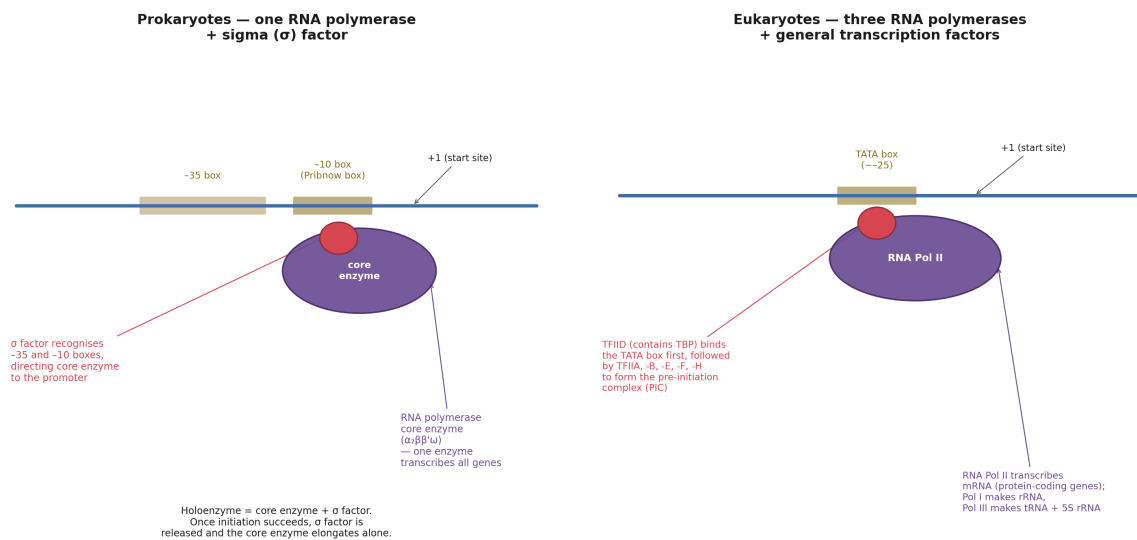


Fig. 2 — Assembly of the transcription initiation complex: a sigma factor guiding the bacterial core enzyme to $-35/-10$ boxes, versus general transcription factors assembling a pre-initiation complex at the eukaryotic TATA box.

(c) Elongation and Termination

- The core enzyme translocates along the DNA, unwinding roughly 17 bp at a time within the transcription bubble, adding ribonucleotides at a rate of about 40–80 nucleotides per second.
- Rho-independent (intrinsic) termination** — the newly made RNA folds into a stable **hairpin (stem-loop)** structure immediately followed by a run of U residues; this combination destabilises the RNA–DNA hybrid and causes the polymerase to dissociate.
- Rho-dependent termination** — the protein **Rho**, an ATP-dependent helicase, binds a specific loading site on the nascent RNA (rut site), tracks along it, and catches up to the stalled polymerase at a pause site, unwinding the RNA–DNA hybrid and releasing the transcript.
- Because bacterial mRNA is used directly for translation, there is generally no post-transcriptional processing of bacterial mRNA (in contrast with the extensive processing of eukaryotic pre-mRNA described in Section 4).

Feature	Prokaryotic transcription
RNA polymerase(s)	One enzyme (core: $\alpha_2\beta\beta'\omega$) transcribes all RNA classes
Promoter elements	-35 box (TTGACA) and -10 /Pribnow box (TATAAT)
Initiation factor	Sigma (σ) factor — confers promoter specificity
Location	Cytoplasm — coupled to translation
Termination	Rho-independent (hairpin + poly-U) or Rho-dependent
Post-transcriptional processing	Minimal — mRNA is generally translated as made

3. Eukaryotic Transcription

Eukaryotic transcription is considerably more elaborate than in bacteria: it takes place inside the nucleus (physically separated from translation, which occurs in the cytoplasm), uses three distinct RNA polymerases, requires numerous accessory transcription factors, and produces transcripts that must be extensively processed before they are functional.

(a) Three RNA Polymerases

- **RNA polymerase I** — located in the nucleolus; transcribes the genes for the large ribosomal RNAs (28S, 18S, and 5.8S rRNA in mammals).
- **RNA polymerase II** — transcribes all protein-coding genes into **pre-mRNA**, as well as many small nuclear and small nucleolar RNAs; it is the polymerase most heavily regulated by transcription factors and is the target of extensive post-transcriptional processing.
- **RNA polymerase III** — transcribes small, stable RNAs: **transfer RNA (tRNA)**, **5S ribosomal RNA**, and other small RNAs (e.g. U6 snRNA).

(b) Initiation at the Core Promoter

- Many RNA polymerase II promoters contain a **TATA box** (consensus TATAAA), typically located about 25–30 bp upstream of the transcription start site, along with other core promoter elements (e.g. the initiator element, downstream promoter element).
- Initiation begins when **TFIID** — via its **TATA-binding protein (TBP)** subunit — binds the TATA box, bending the DNA and nucleating assembly of the remaining **general transcription factors (TFIIA, TFIIB, TFIIE, TFIIF, TFIIH)** together with RNA polymerase II into the **pre-initiation complex (PIC)**.
- **TFIIH** has helicase activity that unwinds the DNA at the start site (forming the open complex) and kinase activity that phosphorylates the **C-terminal domain (CTD)** of RNA polymerase II, a signal that triggers the enzyme's release from the promoter into elongation (promoter clearance).
- Beyond the general transcription factors, gene-specific **activators and repressors** bind regulatory DNA elements (enhancers, silencers), which may lie far from the promoter and act by looping the DNA to contact the initiation complex, often via a large multi-protein **Mediator** complex.

(c) Elongation

- Once past the promoter, RNA polymerase II elongates the transcript with the help of elongation factors that increase its processivity and help it pass nucleosomes.
- Eukaryotic elongation is markedly slower than bacterial elongation (roughly 25–50 nucleotides per second), consistent with the need for extensive co-transcriptional processing (capping and splicing occur even as elongation continues).

(d) Termination

- For RNA polymerase II genes, termination is closely linked to **3' end processing**: a specific **poly-adenylation signal sequence (AAUAAA)** in the nascent RNA is recognised by the **CPSF (Cleavage and Polyadenylation Specificity Factor)** and related proteins, which cleave the RNA at a site downstream of this signal.
- Cleavage exposes a free 5' end on the downstream RNA fragment that is still being made by the polymerase; this uncapped end is degraded by an exonuclease that eventually catches up to and dislodges the polymerase (the "torpedo" model of termination).

Feature	Eukaryotic transcription
RNA polymerases	Three — Pol I (rRNA), Pol II (mRNA + many small RNAs), Pol III (tRNA, 5S rRNA)
Promoter elements	TATA box (~-25), initiator element, downstream promoter element, plus distal enhancers/silencers
Initiation factors	General transcription factors TFIIA/B/D/E/F/H (TFIID contains TBP)
Location	Nucleus — physically separated from translation
Termination (Pol II)	Linked to cleavage at the poly-A signal; "torpedo" exonuclease model
Post-transcriptional processing	Extensive — 5' capping, splicing, 3' polyadenylation (see Section 4)

4. Post-Transcriptional Processing of Eukaryotic mRNA

Unlike bacterial mRNA, the primary transcript made by RNA polymerase II — **pre-mRNA** (also called heterogeneous nuclear RNA, hnRNA) — is not translated directly. It must first undergo three major modifications inside the nucleus before being exported to the cytoplasm as mature mRNA.

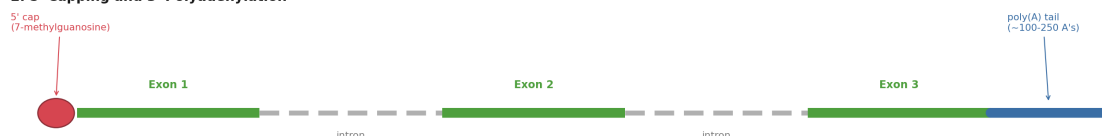
Post-Transcriptional Processing of Eukaryotic Pre-mRNA

1. Primary transcript (pre-mRNA / hnRNA)



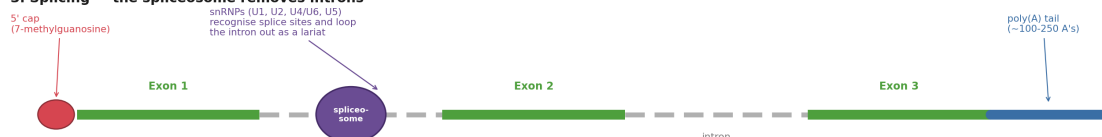
Exons (coding) and introns (non-coding) are both present immediately after transcription.

2. 5' Capping and 3' Polyadenylation



A 7-methylguanosine cap is added co-transcriptionally at the 5' end; a poly(A) tail is added at the 3' end after cleavage at the poly-A signal — both protect the transcript and aid export/translation.

3. Splicing — the spliceosome removes introns



4. Mature mRNA — ready for nuclear export and translation



Introns fully removed; exons spliced directly together, flanked by the 5' cap and poly(A) tail.

Fig. 3 — From primary transcript to mature mRNA: 5' capping and 3' polyadenylation protect the ends, and the spliceosome removes introns so that only exons remain.

(a) 5' Capping

- Almost as soon as the 5' end of the transcript emerges from RNA polymerase II, a **7-methylguanosine (m⁷G) cap** is added via an unusual 5'-to-5' triphosphate linkage.
- The cap protects the transcript from degradation by 5' exonucleases, assists in nuclear export, and is later recognised by initiation factors during translation (cap-dependent ribosome recruitment).

(b) Splicing

- Most eukaryotic genes are interrupted by non-coding sequences called **introns**, which lie between the coding sequences, or **exons**; both are transcribed into the pre-mRNA and introns must be removed before translation.
- Splicing is carried out by the **spliceosome**, a large, dynamic assembly of five **small nuclear ribonucleoproteins (snRNPs: U1, U2, U4, U5, U6)** plus many additional proteins.
- Conserved sequences mark the boundaries of each intron: a **5' splice site (GU)** at its start and a **3' splice site (AG)** at its end, along with an internal branch-point adenosine.
- The spliceosome catalyses two sequential transesterification reactions that join the two flanking exons and release the intron as a looped **lariat** structure, which is subsequently degraded.
- In many genes, **alternative splicing** — the selective inclusion or exclusion of particular exons — allows a single gene to encode multiple distinct protein products, greatly expanding proteome diversity beyond the number of protein-coding genes.

(c) 3' Polyadenylation

- After cleavage at the poly-adenylation signal (described in Section 3d), the enzyme **poly(A) polymerase** adds a chain of roughly 100–250 adenine nucleotides to the free 3' end, forming the **poly(A) tail**.
- The poly(A) tail (bound by poly(A)-binding proteins) protects the 3' end from exonuclease degradation, assists nuclear export, and enhances translational efficiency and mRNA stability in the cytoplasm.

Why so much processing?

Capping, splicing, and polyadenylation are tightly coupled to transcription itself — many of these processing enzymes are recruited directly to the phosphorylated CTD tail of elongating RNA polymerase II, so that a pre-mRNA is capped, spliced, and polyadenylated essentially co-transcriptionally rather than afterward.

5. Quick Revision Summary

Aspect	Key point
Direction of synthesis	RNA synthesised 5'→3'; template strand read 3'→5'; no primer required
Prokaryotic enzyme	Single RNA polymerase (core $\alpha_2\beta\beta'\omega$); sigma factor gives promoter specificity
Prokaryotic promoter	–35 box (TTGACA) and –10/Pribnow box (TATAAT)
Prokaryotic termination	Rho-independent (hairpin + poly-U) or Rho-dependent (Rho helicase)
Eukaryotic enzymes	Pol I (rRNA), Pol II (mRNA), Pol III (tRNA, 5S rRNA)

Aspect	Key point
Eukaryotic promoter	TATA box + core promoter elements, plus distal enhancers/silencers
Eukaryotic initiation factors	TFIIA/B/D/E/F/H assemble the pre-initiation complex; TFIID contains TBP
Eukaryotic mRNA processing	5' m ⁷ G cap, splicing (spliceosome removes introns), 3' poly(A) tail
Key difference	Prokaryotic transcription and translation are coupled; eukaryotic are compartmentalised

Compiled study note — molecular biology reference. Figures are original schematic diagrams prepared for this note and are not to scale.