**Eukaryotic Transcription**

**Initiation of Transcription in Eukaryotes**

Initiation is the first step of eukaryotic transcription and requires RNAP and several transcription factors to proceed.

### Steps in Eukaryotic Transcription

Eukaryotic transcription is carried out in the nucleus of the cell by one of three RNA polymerases, depending on the RNA being transcribed, and proceeds in three sequential stages:

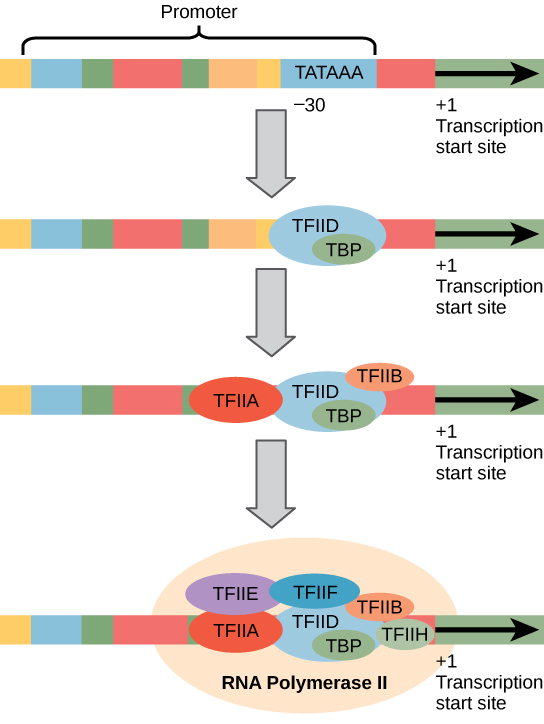
1. Initiation
2. Elongation
3. Termination.

### Initiation of Transcription in Eukaryotes

Unlike the prokaryotic RNA polymerase that can bind to a DNA template on its own, eukaryotes require several other proteins, called transcription factors, to first bind to the promoter region and then help recruit the appropriate polymerase. The completed assembly of transcription factors and RNA polymerase bind to the promoter, forming a transcription pre-initiation complex (PIC).

The most-extensively studied core promoter element in eukaryotes is a short DNA sequence known as a TATA box, found 25-30 base pairs upstream from the start site of transcription. Only about 10-15% of mammalian genes contain TATA boxes, while the rest contain other core promoter elements, but the mechanisms by which transcription is initiated at promoters with TATA boxes is well characterized.

The TATA box, as a core promoter element, is the binding site for a transcription factor known as TATA-binding protein (TBP), which is itself a subunit of another transcription factor: Transcription Factor II D (TFIID). After TFIID binds to the TATA box via the TBP, five more transcription factors and RNA polymerase combine around the TATA box in a series of stages to form a pre-initiation complex. One transcription factor, Transcription Factor II H (TFIIH), is involved in separating opposing strands of double-stranded DNA to provide the RNA Polymerase access to a single-stranded DNA template. However, only a low, or basal, rate of transcription is driven by the pre-initiation complex alone. Other proteins known as activators and repressors, along with any associated coactivators or corepressors, are responsible for modulating transcription rate. Activator proteins increase the transcription rate, and repressor proteins decrease the transcription rate.



**Eukaryotic Transcription Initiation**: 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.

### The Three Eukaryotic RNA Polymerases (RNAPs)

The features of eukaryotic mRNA synthesis are markedly more complex those of prokaryotes. Instead of a single polymerase comprising five subunits, the eukaryotes have three polymerases that are each made up of 10 subunits or more. Each eukaryotic polymerase also requires a distinct set of transcription factors to bring it to the DNA template.

RNA polymerase I is located in the nucleolus, a specialized nuclear substructure in which ribosomal RNA (rRNA) is transcribed, processed, and assembled into ribosomes. The rRNA molecules are considered structural RNAs because they have a cellular role but are not translated into protein. The rRNAs are components of the ribosome and are essential to the process of translation. RNA polymerase I synthesizes all of the rRNAs except for the 5S rRNA molecule.

RNA polymerase II is located in the nucleus and synthesizes all protein-coding nuclear pre-mRNAs. Eukaryotic pre-mRNAs undergo extensive processing after transcription, but before translation. RNA polymerase II is responsible for transcribing the overwhelming majority of eukaryotic genes, including all of the protein-encoding genes which ultimately are translated into proteins and genes for several types of regulatory RNAs, including microRNAs (miRNAs) and long-coding RNAs (lncRNAs).

RNA polymerase III is also located in the nucleus. This polymerase transcribes a variety of structural RNAs that includes the 5S pre-rRNA, transfer pre-RNAs (pre-tRNAs), and small nuclear pre-RNAs. The tRNAs have a critical role in translation: they serve as the adaptor molecules between the mRNA template and the growing polypeptide chain. Small nuclear RNAs have a variety of functions, including “splicing” pre-mRNAs and regulating transcription factors. Not all miRNAs are transcribed by RNA Polymerase II, RNA Polymerase III transcribes some of them.

**Modeling transcription**: This interactive models the process of DNA transcription in a eukaryotic cell.

## Elongation and Termination in Eukaryotes

Elongation synthesizes pre-mRNA in a 5′ to 3′ direction, and termination occurs in response to termination sequences and signals.

### Transcription through Nucleosomes

Following the formation of the pre-initiation complex, the polymerase is released from the other transcription factors, and elongation is allowed to proceed with the polymerase synthesizing RNA in the 5′ to 3′ direction. RNA Polymerase II (RNAPII) transcribes the major share of eukaryotic genes, so this section will mainly focus on how this specific polymerase accomplishes elongation and termination.

Although the enzymatic process of elongation is essentially the same in eukaryotes and prokaryotes, the eukaryotic DNA template is more complex. When eukaryotic cells are not dividing, their genes exist as a diffuse, but still extensively packaged and compacted mass of DNA and proteins called chromatin. The DNA is tightly packaged around charged histone proteins at repeated intervals. These DNA–histone complexes, collectively called nucleosomes, are regularly spaced and include 146 nucleotides of DNA wound twice around the eight histones in a nucleosome like thread around a spool.

For polynucleotide synthesis to occur, the transcription machinery needs to move histones out of the way every time it encounters a nucleosome. This is accomplished by a special protein dimer called FACT, which stands for “facilitates chromatin transcription.” FACT partially disassembles the nucleosome immediately ahead (upstream) of a transcribing RNA Polymerase II by removing two of the eight histones (a single dimer of H2A and H2B histones is removed.) This presumably sufficiently loosens the DNA wrapped around that nucleosome so that RNA Polymerase II can transcribe through it. FACT reassembles the nucleosome behind the RNA Polymerase II by returning the missing histones to it. RNA Polymerase II will continue to elongate the newly-synthesized RNA until transcription terminates.

### Elongation

RNA Polymerase II is a complex of 12 protein subunits. Specific subunits within the protein allow RNA Polymerase II to act as its own helicase, sliding clamp, single-stranded DNA binding protein, as well as carry out other functions. Consequently, RNA Polymerase II does not need as many accessory proteins to catalyze the synthesis of new RNA strands during transcription elongation as DNA Polymerase does to catalyze the synthesis of new DNA strands during replication elongation.

However, RNA Polymerase II does need a large collection of accessory proteins to initiate transcription at gene promoters, but once the double-stranded DNA in the transcription start region has been unwound, the RNA Polymerase II has been positioned at the +1 initiation nucleotide, and has started catalyzing new RNA strand synthesis, RNA Polymerase II clears or “escapes” the promoter region and leaves most of the transcription initiation proteins behind.

All RNA Polymerases travel along the template DNA strand in the 3′ to 5′ direction and catalyze the synthesis of new RNA strands in the 5′ to 3′ direction, adding new nucleotides to the 3′ end of the growing RNA strand.

RNA Polymerases unwind the double stranded DNA ahead of them and allow the unwound DNA behind them to rewind. As a result, RNA strand synthesis occurs in a transcription bubble of about 25 unwound DNA basebairs. Only about 8 nucleotides of newly-synthesized RNA remain basepaired to the template DNA. The rest of the RNA molecules falls off the template to allow the DNA behind it to rewind.

RNA Polymerases use the DNA strand below them as a template to direct which nucleotide to add to the 3′ end of the growing RNA strand at each point in the sequence. The RNA Polymerase travels along the template DNA one nucleotide at at time. Whichever RNA nucleotide is capable of basepairing to the template nucleotide below the RNA Polymerase is the next nucleotide to be added. Once the addition of a new nucleotide to the 3′ end of the growing strand has been catalyzed, the RNA Polymerase moves to the next DNA nucleotide on the template below it. This process continues until transcription termination occurs.

### Termination

The termination of transcription is different for the three different eukaryotic RNA polymerases.

The ribosomal rRNA genes transcribed by RNA Polymerase I contain a specific sequence of basepairs (11 bp long in humans; 18 bp in mice) that is recognized by a termination protein called TTF-1 (Transcription Termination Factor for RNA Polymerase I.) This protein binds the DNA at its recognition sequence and blocks further transcription, causing the RNA Polymerase I to disengage from the template DNA strand and to release its newly-synthesized RNA.

The protein-encoding, structural RNA, and regulatory RNA genes transcribed by RNA Polymerse II lack any specific signals or sequences that direct RNA Polymerase II to terminate at specific locations. RNA Polymerase II can continue to transcribe RNA anywhere from a few bp to thousands of bp past the actual end of the gene. However, the transcript is cleaved at an internal site before RNA Polymerase II finishes transcribing. This releases the upstream portion of the transcript, which will serve as the initial RNA prior to further processing (the pre-mRNA in the case of protein-encoding genes.) This cleavage site is considered the “end” of the gene. The remainder of the transcript is digested by a 5′-exonuclease (called Xrn2 in humans) while it is still being transcribed by the RNA Polymerase II. When the 5′-exonulease “catches up” to RNA Polymerase II by digesting away all the overhanging RNA, it helps disengage the polymerase from its DNA template strand, finally terminating that round of transcription.

In the case of protein-encoding genes, the cleavage site which determines the “end” of the emerging pre-mRNA occurs between an upstream AAUAAA sequence and a downstream GU-rich sequence separated by about 40-60 nucleotides in the emerging RNA. Once both of these sequences have been transcribed, a protein called CPSF in humans binds the AAUAAA sequence and a protein called CstF in humans binds the GU-rich sequence. These two proteins form the base of a complicated protein complex that forms in this region before CPSF cleaves the nascent pre-mRNA at a site 10-30 nucleotides downstream from the AAUAAA site. The Poly(A) Polymerase enzyme which catalyzes the addition of a 3′ poly-A tail on the pre-mRNA is part of the complex that forms with CPSF and CstF.

The tRNA, 5S rRNA, and structural RNAs genes transcribed by RNA Polymerase III have a not-entirely-understood termination signal. The RNAs transcribed by RNA Polymerase III have a short stretch of four to seven U’s at their 3′ end. This somehow triggers RNA Polymerase III to both release the nascent RNA and disengage from the template DNA strand.