Molecular Biology
Lecture 6

Transcription (Transcription of DNA to RNA)

Transcription is the biochemical process of transferring the information in a DNA sequence to an RNA molecule. The RNA molecule can be the final product, or in the case of messenger RNA (mRNA), it can be used in the process of translation to produce proteins. RNA Polymerase is a protein complex that performs the main job of reading a DNA template and synthesizing RNA, but accessory proteins are also needed.

The enzyme used in transcription is "**RNA polymerase**". There are several forms of RNA polymerase. Most genes are transcribed by **RNA polymerase II**.

The raw materials for the new RNA are the 4 ribonucleoside triphosphates: **ATP, CTP, GTP, and UTP**.

As with DNA replication, transcription proceeds 5′ to 3' direction : new bases are added to the free 3' OH group.

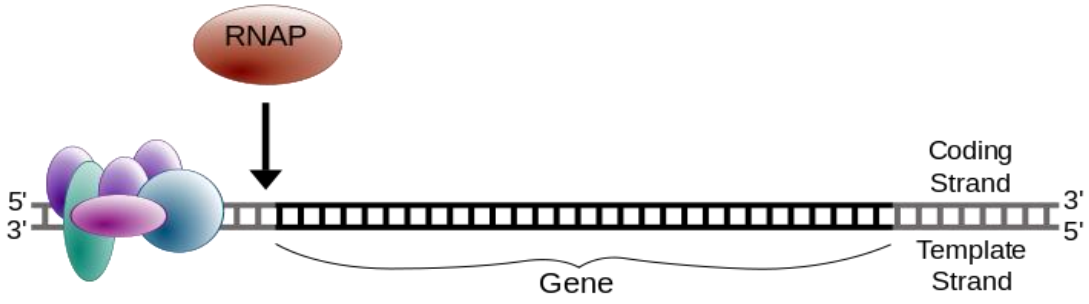
Unlike replication, transcription does not need to build on a primer. Instead, transcription starts at a region of DNA called a "**promoter**". For protein-coding genes, the promoter is located a few bases 5' to (upstream from) the first base that is transcribed into RNA. Promoter sequences are very similar to each other, but not identical.

Process of Transcription

Transcription is divided into 5 steps :

Pre-initiation

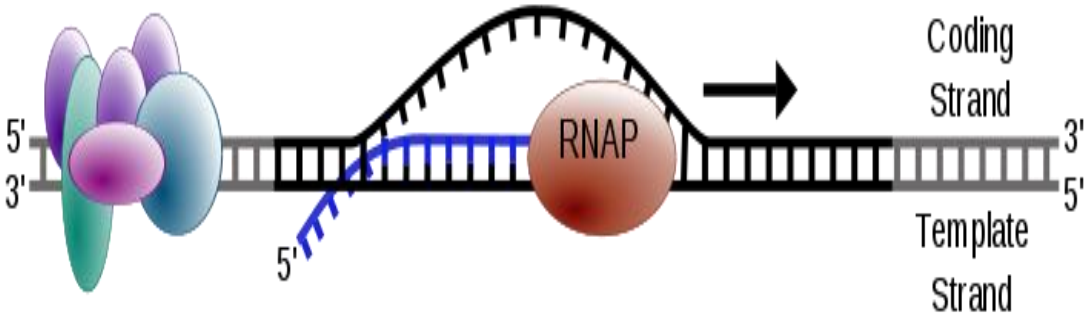
RNA polymerase requires the presence of a core promoter sequence in the DNA. Promoters are regions of DNA which promote transcription. RNA polymerase is able to bind to core promoters in the presence of various specific transcription factors.The most common type of core promoter in eukaryotes is a short DNA sequence known as a TATA box, found -30 base pairs from the start site of transcription. The TATA box, as a core promoter, is the binding site for a transcription factor known as TATA binding protein (TBP), which is itself a subunit of another transcription factor, five more transcription factors and RNA polymerase combine around the TATA box in a series of stages to form a preinitiation complex. One transcription factor, DNA helicase, has helicase activity and so is involved in the separating of opposing strands of double-stranded DNA to provide access to a single-stranded DNA template.



Initiation
The completed assembly of transcription factors and RNA polymerase bind to the promoter, forming a transcription initiation complex.

Promoter clearance
After the first bond is synthesized, the RNA polymerase must clear the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts.

Elongation
One strand of the DNA, the *template strand* (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy. Although RNA polymerase traverses the template strand from 3' → 5', the coding (non-template) strand and newly-formed RNA can also be used as reference points, so transcription can be described as occurring 5' → 3'. This produces an RNA molecule from 5' → 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone). Unlike DNA replication, mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene. Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases.



Termination
A protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.

After Transcription
In prokaryotes, the RNA copy of a gene is messenger RNA, ready to be translated into protein.

In eukaryotes, the primary RNA transcript of a gene needs further processing before it can be translated. This step is called "RNA processing". Also, it needs to be transported out of the nucleus into the cytoplasm.

- Steps in RNA processing:
  - 1. Add a cap to the 5' end
  - 2. Add a poly-A tail to the 3' end
  - 3. splice out introns.

Capping (Add a cap to the 5' end**)**
RNA is unstable, especially at the ends. The ends are modified to protect it. At the 5' end, a slightly modified guanine is attached "backwards", by a 5' to 5' linkage, to the triphosphates of the first transcribed base.

Polyadenelation
At the 3' end, the primary transcript RNA is cut at a specific site and **100-200 adenine nucleotides** are attached, the **poly-A tail**. Note that these A's are not coded in the DNA of the gene.

Splice out introns.
Introns are regions within a gene that don't code for protein and don't appear in the final mRNA molecule. Protein-coding sections of a gene (called exons) are interrupted by introns.
The function of introns remains unclear. They may help is RNA transport or in control of gene expression in some cases.
Some genes have many long introns: the dystrophin gene (mutants cause muscular dystrophy) has more than 70 introns that make up more than 99% of the gene's sequence. However, not all eukaryotic genes have introns: histone genes, for example, lack introns.

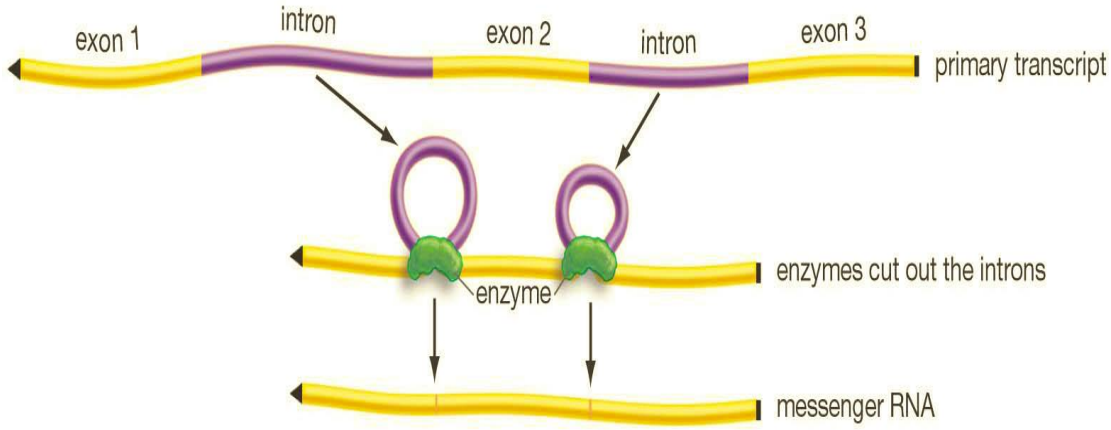


Figure : Introns are "spliced out" by RNA/protein hybrids called "spliceosomes". The intron sequences are removed, and the remaining ends are re-attached so the final RNA consists of exons only.