

# MORE TRANSCRIPTION - A CLOSER LOOK\*

Andrew Hughes

This work is produced by The Connexions Project and licensed under the  
Creative Commons Attribution License †

## 1 Properties of Eukaryotic Transcription

- Complex!
- RNA polymerase is the main protein that creates the RNA transcript
- Proteins called transcription factors help RNA polymerase and regulate it's function.
- 2 types of transcription factors: basal tfs and modulatory tfs.
- Basal transcription factors - required for all transcription events.
- Modulatory transcription factors - regulate expression of genes. Both positive and negative regulation.
- Transcription factors can act from relatively far away. Up to 10kb!
- pre-mRNA transcripts are spliced to remove exons.
- Alternative splicing allows the cell to combine exons differently in the final mRNA transcript from the same pre-mRNA transcript.

## 2 Transcription In Five Easy Steps

1. Transcription of DNA to pre-mRNA. DNA, RNA polymerase and transcription factors.
2. Addition of 5' methyl guanosine cap to pre-mRNA transcript.
3. Splicing of pre-mRNA transcript (yields mRNA proper)
4. Addition of 3' poly(A) tail to mRNA transcript.
5. mRNA is transported to the cytoplasm.

---

\*Version 1.2: Jul 27, 2003 5:35 pm GMT-5

†<http://creativecommons.org/licenses/by/1.0>

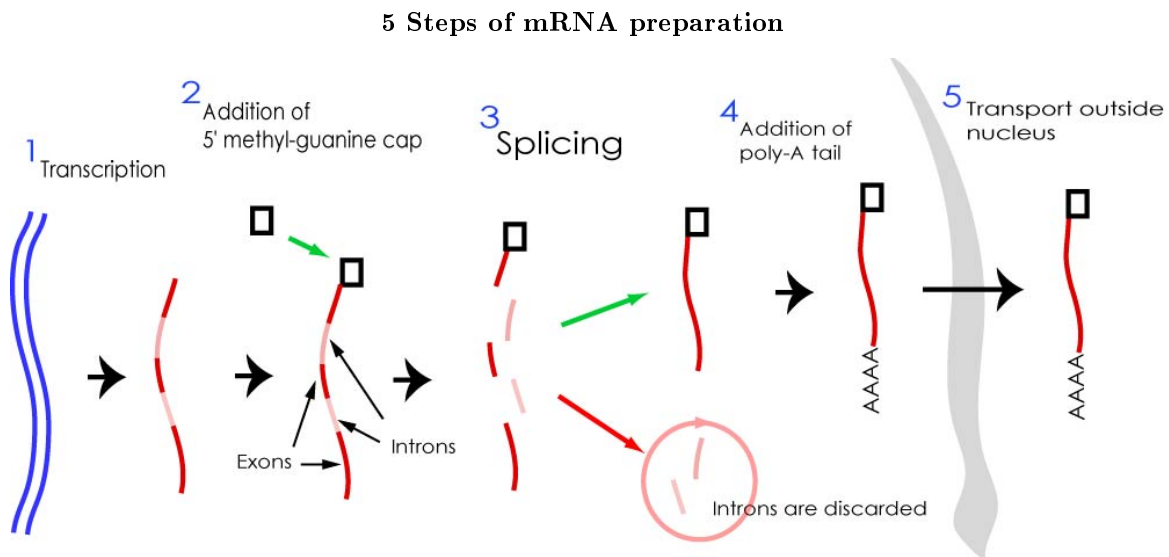


Figure 1

### 3 RNA Polymerase - main protein that creates RNA transcript.

RNA polymerase is the protein whose job it is to 'read' the genetic code and create a complimentary RNA transcript from that code. Eukaryotes have three different types of RNA polymerases: RNA polymerase I, II, and III. RNA polymerase II is the form of polymerase that transcribes most genes and is the form of polymerase with which we need to concern ourselves.

### 4 Initiation of Transcription

Transcription initiation in eukaryotes is complicated, and the details are not entirely understood (although we do have a good grasp on the basic mechanism). The fully assembled eukaryotic transcription initiation complex contains more than 50 polypeptides. RNA polymerase II has more than 10 polypeptide subunits by itself. Keep in mind that the process outlined below is a generalization, and any given, specific transcription event will probably vary some in the details. For example, although the TATA box is the most strongly conserved promoter sequence, it is by no means present in every eukaryotic promoter. Also, there is some debate as the importance of the order of the binding of the proteins in the initiation complex: it was thought that a specific order was vital, now, however, new evidence suggests that reaching the end binding state may be what is really important, whatever the order.

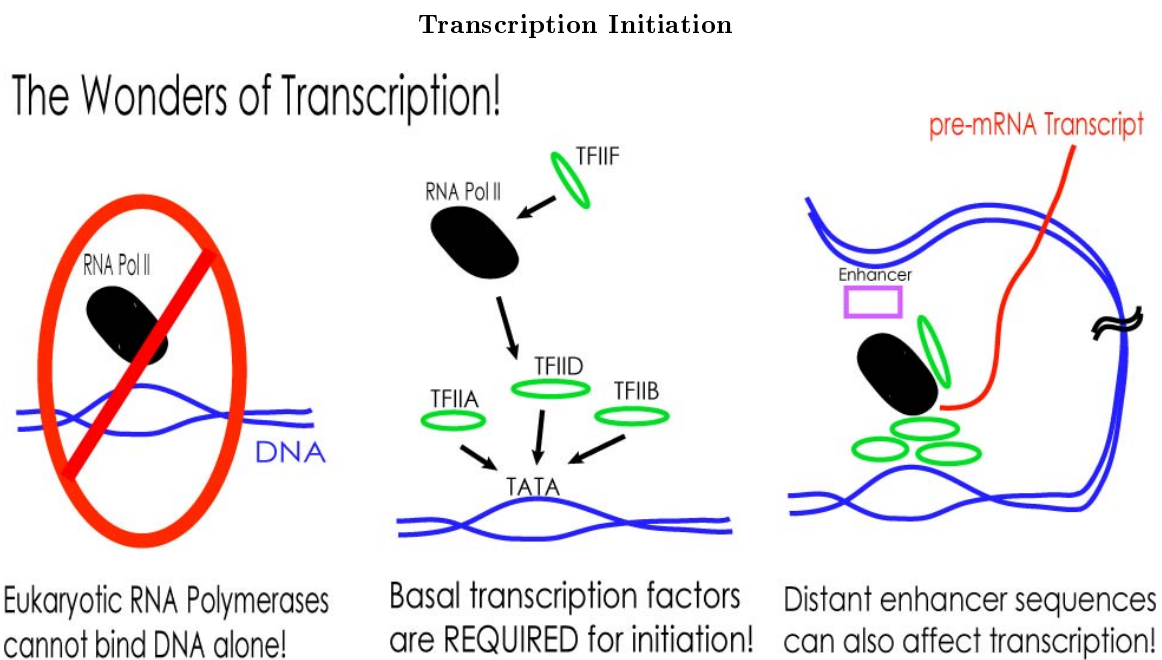
#### Transcription Factors - help/regulate RNA polymerase's function

Transcription factors (tfs) are proteins involved with transcription - except for RNA polymerase. Transcription factors can be broken down into two groups: basal tfs and modulatory tfs. Eukaryotic RNA polymerases are not capable of initiating transcription alone, they require the assistance of a set of basal transcription factors. Basal tfs assist RNA polymerase in the recognition of promoter sequences and unwinding the DNA double helix, among other functions. Basal tfs are necessary for every transcription event. Modulatory transcription factors regulate the expression of a gene, or a set of gene. These tfs are important because they

allow the body to differentially express genes at different times and different places in the body. Modulatory tfs are vital for multicellular life. They allow the body to create different cells, tissues, and organs.

- Learning the specific order of tf binding is not vital
- Understand that transcription initiation involves MANY polypeptides, each with a specific function.
- Functions include: recognizing promoter sequences, modulating amount of transcript production, modulating time/space of transcript production, and unwinding DNA
- Each tf/protein represents a functional unit

Transcription factors are denoted TFIIX, for Transcription Factor for polymerase II. X is an identifying letter.



**Basal Transcription Factors - necessary for transcription**

The first step in transcription initiation is TFIIA binds to the TATA box in the promoter region. The TATA box is the most strongly conserved consensus sequence in the eukaryotic promoter; it has the consensus sequence TATAAAA and is usually located around position -30 (counting backwards from the transcriptional start site). After TFIID binds to the TATA box, TFIIA joins the initiation complex, followed by TFIIB. Separate from the initiation complex, TFIIF binds to RNA polymerase II. One subunit of TFIIF has DNA-unwinding properties that presumably help RNA polymerase II unwind the DNA during transcription. Once the TFIIF-RNA polymerase complex has been formed, the two proteins can then join the initiation complex. After the addition of several more proteins to the initiation complex, transcription is finally ready to begin.

**Steps of Transcription Initiation - Basal Transcription Factors**

1. TFIID binds to the TATA box in the promoter region.

2. TFIIA joins initiation complex.
3. TFIIB joins initiation complex.
4. TFIIF binds RNA polymerase (separate from initiation complex).
5. RNA poly - TFIIF complex binds initiation complex on DNA.
6. Transcription begins.

### **Modulatory Transcription Factors - regulate transcription**

Regulatory transcription factors are non-intrinsic transcription factors that modulate the expression of a particular transcript. These are the transcription factors that regulate time/space differential expression in the organism. These factors can be enhancers or silencers, meaning that they can both increase and decrease the expression of a given gene. Also, these transcription factors can act from several thousand base-pairs in either direction from the promoter. Here it is important to remember that the eukaryotic genome has significant secondary structure; DNA wrapping around nucleosomes and 30nm fiber looping can bring distant sequence motifs proximal to promoter sites where they can have significant action. It is important to remember that an enhancer/silencer sequence motif may be present without having an effect, because the specific transcription factor protein responsible for binding to that domain may not have been expressed, i.e. may not be present. The sequence on its own has no effect, the presence of the modulatory transcription factor is required. This can complicate sequence analysis w/r/t expression because in some cases we will only have half of the contextual information (ie we are missing the protein context of the cell).

## **5 Transcript Modifications**

### **5.1 5' Methyl Guanosine Cap**

- Protects transcript from degradation.
- Helps with initiation of translation.

After the initial pre-mRNA transcript has been created the first major modification is the addition of a 5' methyl guanosine cap. The addition of the 5' 7-MG cap is important for two reasons: the 5' caps are recognized by protein factors that initiate translation, and it also helps protect the transcript from nucleases. Nucleases are very common in the cell and because of this unprotected RNA has a very short half-life inside the cell. Nucleases are actually so common that working with RNA in the laboratory can be quite difficult because the samples have a tendency to disintegrate into useless bits.

### **5.2 Splicing**

- Splicing removes non-coding exons from transcript.
- Alternative splicing allows for different combinations of exons from same pre-mRNA transcript (gene).
- Some RNAs can self-splice.

Eukaryotic genes contain two types of transcribed regions: introns and exons. Exons are the regions of the genome that contain actual coding information. Introns are non-coding, meaning that intronic sequences are never translated to protein. Introns are never included in the final processed mRNA transcript. Splicing is the process of removing introns from the pre-mRNA transcript to produce an exon-only mRNA molecule, which is then shipped off for translation. Generally, eukaryotic mRNAs are considered to be monogenic. Monogenic means that an RNA transcript contains exons from only one gene. However, up to one fourth of the transcripts in *C. elegans* have been shown to be multi-genic (i.e. they contain exons from multiple genes).

A further complication of the splicing process is that mRNA can undergo alternative splicing. To illustrate this let's imagine a gene that has 3 exons and two introns. From this gene, three different final transcripts are possible. In all transcripts the two introns are going to be removed, but the cell can combine the exons however it wants as long as the original order is maintained. This means that for this example the possible mRNA transcripts include: Exon1-Exon2, Exon1-Exon3, and Exon1-Exon2-Exon3; however, Exon3-Exon1 is not possible because the exons are out of order.

An interesting side note is that some introns are capable of self-splicing, that is they can politely remove themselves without the intervention of any proteins. This is significant mainly because it is a significant counter example to the idea that RNA is an inert transcript and action is solely the domain of proteins. RNAs should really be viewed as having both enzymatic properties and abstract information-carrying ability. Because of this many people believe that RNA was the original genetic molecule and that DNA and proteins evolved later in the game.

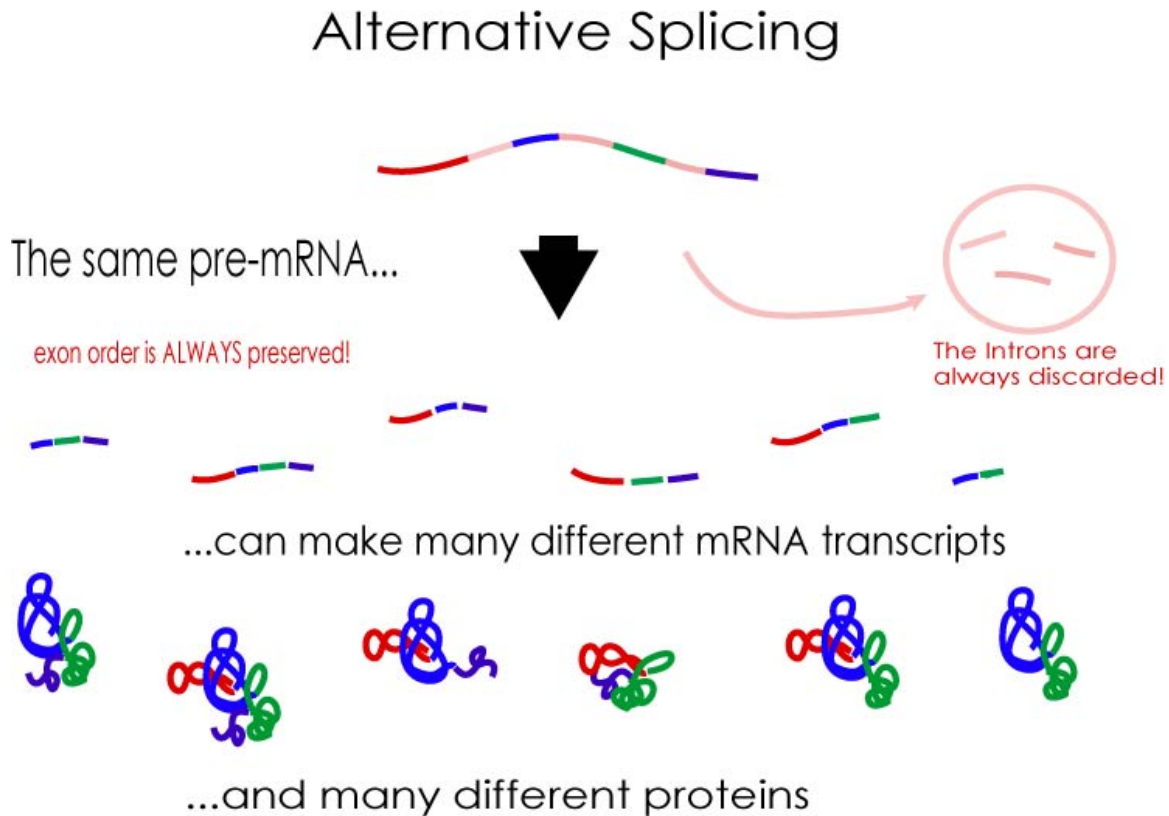


Figure 3

Alternative splicing is a very important and powerful tool. To understand the benefit alternative splicing gives the cell we need to understand something about proteins. Proteins can be understood as containing modularized functional units. These functional units can be active sites on enzymes, large structural motifs such as beta-sheets or alpha-helices, or motifs that direct the eventual destination of expressed proteins. A good example of an alternatively spliced pre-mRNA transcript is the mouse IgM immunoglobulin transcript. IgM exists in two forms: secreted and membrane bound. These two forms of the protein differ in the only in the C-terminus: the secreted protein has a secreted terminus motif while the membrane-bound protein has a C-terminal membrane anchor region. Both products come from the same pre-mRNA, but alternative splicing includes either the terminal exon that creates the secreted form of IgM or the membrane-bound form of IgM.

### 5.3 3' poly-Adenylation

- Important for cellular transport.
- Helps stabilize the transcript

The poly(A) tails are formed in a two step process: an endonuclease cleaves around 1000-2000 non-coding bases from the 3' end of the pre-mRNA transcript and then poly(A) polymerase adds 20-200 AMP molecules to the 3' end of the transcript. The poly(A) tail is important in the cellular transport of the mRNA transcript and, like the 5' cap, also helps to stabilize the mRNA transcript.