Transcription- lecture outline

For most organisms genetic information is stored in DNA

The need for a "messenger" for conveying and using the stored information, that is, gene expression, which generally means protein synthesis, was made apparent by:

- in eukaryotes, almost all the DNA is in the nucleus, but the use of labeled amino acids showed protein synthesis occurs in the cytoplasm on ribosomes.

- red blood cells of animals make protein long after the nucleus is gone.

- when the stalk or cap of a unicellular alga was cut off, even the part without a nucleus could restore the missing part (but could not reproduce).

- the possibility that each ribosome contained a unique set of information was eliminated by "swapping" ribosomes.

These observations led to the hypothesis that a message must be involved and based on the appearance of RNA very soon after infection of \textit{E. coli} with DNA phage, RNA seemed the most logical candidate. See \url{http://www.ornl.gov/info/ornlreview/v37_3_04/article12.shtml} for a review of the discovery, or for an intro level animation showing the people involved, go to: \url{http://www.dnaftb.org/dnaftb/21/animationanimation.html}

mRNA is a single stranded molecule that typically has a very short half-life, in part because there are many very stable RNAases in most cells

The process of making RNA from DNA is called \textit{transcription};

Transcription requires the enzyme DNA-dependent RNA polymerase. The enzyme uses one strand, the template or antisense strand of DNA, to align and connect complementary RNA nucleotides in a 5' to 3' direction.
Transcription begins when the RNA polymerase attaches to a **promoter**, a region (generally) preceding (upstream) of the coding sequence.

In **prokaryotes**, all classes of RNA are transcribed by the same enzyme so the promoters share common features

- a **TATAAT** "box" about -10 bases from the start of transcription (or Pribnow box)
- a **TTGACA** sequence centered at -35

These are "consensus" sequences, meaning not all are identical, but each box may differ by a few bases.

A sigma factor (generally $\sigma_{70}$) is required to bind the RNA polymerase ($\alpha_{2}\beta'$) to the promoter

- there are sets of genes that have somewhat different promoters and these use different sigma factors. Examples include "heat shock" genes, nitrogen starvation genes, and genes that function in spore formation.

- the average rate of transcription is about 1,000 nucleotides per minute. The Nus A protein in *E. coli* can bind to the core polymerase and slow it down so that transcription does not get too far ahead of translation.

Transcription also ends with built in signals. In *E. coli*, an inverted repeat followed by TAAAAAAA acts as an intrinsic "end transcripton" signal while other somewhat conserved sequences are bound by "rho" protein to prevent further transcription. In eukaryotes, AAUAAA in the appropriate context ends transcription 11-30 bases downstream and also acts as a signal for addition of a 50-250 A tail by a polyA polymerase.

**Transcription in Eukaryotes:**

There are 3 different RNA polymerases that use different promoters

RNApol I reads "large" RNA transcripts, including the multiple copies of rRNA that are made as one long RNA and then cut by specific enzymes
into subunits (prokaryotes also make one long pre-rRNA that is then cut into fragments. *E. coli* has 7 copies of rDNA.

**RNApol II** is used to transcribe mRNAs

**RNApol III** reads short transcripts including tRNA

The promoter signals for pol II are similar to those for prokaryotic promoters except the TATA box is at -25 and there is a **CAAT** box at -75

There are many other small conserved sequences in the upstream region that may be called response elements, enhancer elements or silencers. These "**cis-elements**" are bound by "**trans-acting factors**" to increase or decrease transcription from any promoter. There are often multiple interacting factors that affect the level of expression.

**mRNA processing**

Almost all eukaryotic genes are much longer than the final mRNA encoded. The gene is transcribed from the promoter to the terminator, which leads to addition of a poly A tail in a "pre-mRNA". Coding sequences included in the mRNA are called **exons** (expressed) and the non-coding segments are called **introns** for intervening sequences.

An inverted G is added to the 5' base (almost always A) in a 5'-5' linkage. This "cap" helps protect the mRNA from 5' exonucleases.

**Introns** are removed and **exons** connected to create the final message in a process called "**splicing**". Splicing requires a number of snRNPs, which include both RNA and protein components. Each intron begins with UG and ends with AG, with adjacent (internal) conserved consensus sequences. snRNPs attach to the intron, form a lariat, clip it out and join the flanking exons. The removal occurs in a preferred order. Plant introns are generally shorter than animal introns.


Introns in mitochondrial genes are often removed using "self-splicing".
Extra bases in tRNAs and rRNAs are clipped using specific enzymes.

Many enzymatic alterations are made to bases in tRNA and rRNA.