**Transcription-Outline**

Genes are made of DNA, almost all of which is in the chromosomes.

In **eukaryotes** (organisms which we generally call higher organisms), each cell has a "nucleus". (In Greek, eu is true and karyon is nucleus, so eukaryotes have true nuclei) In eukaryotes, chromosomes are found inside the nucleus. However, protein synthesis in eukaryotes occurs on ribosomes in the cytoplasm, not in the nucleus.

Cells that lose their nuclei, like red blood cells in humans, can still make protein for some time in the absence of DNA.

These kinds of observations led to the conclusion that there must be some sort of a messenger to convey genetic information from genes to the cytoplasm.

The messages, called **mRNA**, are single stranded polymers made using RNA nucleotides.

The classical flow of genetic information, often referred to as "The Central Dogma" is:

DNA ➔ mRNA ➔ protein

The process of making mRNA is called **transcription**. The process is very similar in eukaryotes and prokaryotes (organisms where the DNA is not contained in a nucleus), although differences are found.

<table>
<thead>
<tr>
<th><strong>Eukaryotes</strong></th>
<th><strong>Prokaryotes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>higher organisms</td>
<td>bacteria and bluegreen algae</td>
</tr>
<tr>
<td>nuclear membrane present</td>
<td>no nucleus as such</td>
</tr>
<tr>
<td>DNA in chromosomes</td>
<td>naked DNA</td>
</tr>
<tr>
<td>mitosis and meiosis</td>
<td>no mitotic apparatus</td>
</tr>
<tr>
<td>3 different RNA polymerases</td>
<td>one RNA polymerase</td>
</tr>
<tr>
<td>POL I - large RNAs</td>
<td></td>
</tr>
<tr>
<td>POLII - messages</td>
<td></td>
</tr>
<tr>
<td>POL III small RNAs</td>
<td></td>
</tr>
</tbody>
</table>
In order to transcribe a gene:

Base sequences in the "upstream" region of a gene are recognized and bound by RNA polymerases; this binding region is called a promoter.

Promoters share common "motifs"

   In prokaryotes, there is a TATAAT (TATA-box) approximately 10 bases before the start point of transcription ("-10" in genetic parlance). There is similar sequence in eukaryotic genes that code for proteins (transcribed by POL II) at approx. -25.

   Another sequence, that generally includes a CAAT (CAT-box) is found at -35 in prokaryotes and -75 in eukaryotes.

   (These motifs are called consensus sequences, and individual promoter sequences may vary slightly; this results in some promoters being more effectively transcribed than others.)

RNA polymerase is a large molecule with 5 polypeptide subunits and requires another protein called a sigma factor to attach to a promoter.

RNA polymerase moves down the DNA double helix, unwinds it and uses one of the two strands as a template for making a complementary strand of mRNA as it goes.

   The information on which strand to copy comes from the promoter

   The TATA box signals the start point for transcription

   The direction of synthesis is always 5’ to 3’

   RNAs are made as single stranded molecules

   Messages are fragile; the half life in bacteria is < 3 min.
Transcription ends when the RNA polymerase hits a "built-in" termination signal (often AAAAAAT in the template strand) or in prokaryotes, some stops occur when a protein is bound so tightly transcription can't proceed. (These are called rho-stops)

Special features of eukaryotic messages:

The termination signal triggers another enzyme to add a "poly-A tail, a long stretch of A's to the 3' end of almost all messages.

An inverted G is added to the 5' beginning of the messages ("5' cap")

These additions help to protect messages from exonucleases, enzymes that nibble away the ends of RNA polymers.

Eukaryotic messages must be processed to eliminate "introns".

Eukaryotic genes contain coding sequences for making protein (called exons for expressed segments) interspersed with DNA that is noncoding (introns) that must be removed to make a useful message.

see http://www.swbic.org/education/comp-bio/images/SplicingMechanism.gif

A pre-RNA (also called hnRNA for heterogenous-nuclear RNA) is made that includes a copy of the entire gene, and then the introns are removed using SNRPS (small nuclear ribonuclear/proteins) to cut them out. This is often referred to as splicing since the exons flanking the removed intron are spliced together. Introns begin with a UG and end with AG, but these sequences also appear at non-splice sites! In many cases, more of the DNA sequence in a gene is part of introns than the part that ends up in actual coding sequence.

(The human gene that is defective in Muscular Dystrophy is over 2.4 million bases long, and has over 100 introns.)

In prokaryotes, genes and messages are co-linear, and translation can begin as soon as the 5' beginning of the message has been made.