Transcription

From DNA to RNA

mRNA

In order to get the information stored in the DNA to the site of protein synthesis, a copy of the information has to be made. This is done by using bases on the sense strand of the DNA as a template for synthesizing a complementary strand of RNA.

RNA synthesis

Just as DNA synthesis uses a DNA polymerase, so RNA synthesis uses an RNA polymerase. This enzyme is capable of linking together RNA nucleotides by the same 3'→5' bond as seen in DNA. It will only add a base if it forms stable hydrogen bonds with the DNA base on the other strand.
Prokaryotes vs eukaryotes

There are notable differences in transcription between the two.

Prokaryotes use a single polymerase for all transcription. Eukaryotes use three different polymerases.

Prokaryotes do not need to modify their mRNA before use. Eukaryotes make many changes to the mRNA.

Prokaryotes can use the mRNA for protein synthesis even while it is being made. Eukaryotes must export the mRNA to the cytoplasm.

Prokaryotic enzyme structure

The core enzyme contains two identical α subunits (36kD) and two closely related β (150kD) and β'(155kD) subunits. The α is used in enzyme assembly and in RNA synthesis. The β is needed for chain initiation and elongation while β' is required for binding to the DNA template.

Lack of specificity

If the core enzyme is added to a tube containing DNA and the four nucleotides, it will make RNA. However, the molecules are of random length and do not start at the beginning of a gene, but anywhere along the strand. Clearly this is not what is wanted. More detailed studies of the enzyme showed that, at the time of attachment to DNA there was yet another subunit.
**Sigma (σ) factors**

Sigma (70kD) binds to the core enzyme to form the functional holoenzyme. In this form the enzyme recognizes where a gene starts and attaches to the DNA. At this point the sigma protein leaves. There are many different sigma proteins and each has a preference for specific genes.

**Where to start**

Since a gene contains a string of bases that code for a particular protein (the coding sequence), clearly you want to transcribe all of this information. That suggests that the polymerase will want to attach a little before the start of the coding sequence.

**Locating bases relative to a gene**

By convention, the coding bases of a gene are numbered -1, -2 etc. Those before the coding sequence are numbered -1, -2 etc. Bases before the coding sequence are said to be 'upstream'; those in or after the coding sequence are 'downstream'.
1. Initiation begins
Sigma binds to promoter region of DNA.

2. Initiation continues
Sigma opens the DNA helix; transcription begins.

3. Initiation is complete
Sigma releases; mRNA synthesis continues.
**Promoter structure**

When several promoters are studied, it is clear they all have similar base sequences. While there are some differences, the **consensus sequence** is TTGACA. This means that a T is in first place more often than any other etc. It is this sequence that the holoenzyme binds to. In general, the closer a sequence is to the consensus, the stronger the binding of the polymerase.

**Reading frames**

Since the code is read as a continuous stream of codons, with no spaces between ‘words’, it is essential that the polymerase starts at the right place. Note that, by starting at the wrong bases, all the information will be wrong.

- CGTTATGCACTGACGATACGGATCAGG
- GTTATGCAGTACGACGATACGGATCAGG
- TTATGCACTGACGATACGGATCAGG

**Additional binding**

In addition to the -35 sequence there is a second binding site at -10. This has a **consensus sequence** of TATAAT and is usually referred to as the Pribnow box.
**The complete promoter**

The entire sequence contains the -35 site, 16-18 bases an then the Pribnow box. Changing any of these severely reduces binding. While this is where the enzyme binds, it covers a stretch of DNA from -55 to +20 bases.

TTGACA ---- 16-18---- TATAAT

**Initiation of RNA synthesis**

Once the enzyme binds, and sigma has left, it separates (melts) the two strands of DNA over a short (~17bp) stretch. The enzyme starts to move down the DNA and when reaches the +1 base, it starts to incorporate RNA bases. Often there are several false starts in which a short stretch of RNA is discarded and the process restarts.

**Elongation**

The elongation phase involves the actual synthesis of RNA. The enzyme adds nucleotides to the growing RNA strand, forming a phosphodiester bond. Hydrogen bonds form between complementary base pairs, ensuring accuracy in the synthesis.
Error correction

As you might expect, RNA polymerase needs to be accurate in its copying of genetic information. The active site is designed to be able to remove nucleotides as well as add them to the growing strand. The enzyme tends to hover around mismatched nucleotides longer than properly added ones, giving the enzyme time to remove them. This process is somewhat wasteful, since proper nucleotides are also occasionally removed, but this is a small price to pay for creating better RNA transcripts. Overall, RNA polymerase makes an error about once in 10,000 nucleotides added, or about once per RNA strand created.

Termination

Once the polymerase has transcribed the entire coding region, it is time to release the mRNA. This happens at the 'termination sequence'. While there are variations in the nature of the sequence, they all rely on slowing down or stopping the polymerase. If the polymerase slows down, its binding to the DNA is not enough to maintain a stable structure. As a result it will detach from the DNA and at the same time release the mRNA.

Termination

Most termination sequences have a GC rich region, followed by an AT rich area. The GC region shows one of the remarkable abilities of RNA. Since it is single-stranded, it can fold back on itself and form a local double-stranded structure.
**How Transcription Ends**

1. RNA polymerase reaches a transcription termination signal, which codes for RNA that forms a hairpin.
2. The RNA hairpin causes the RNA strand to separate from the RNA polymerase, terminating transcription.

**Eukaryotic Transcription**

Eukaryotes have three different RNA polymerases. Each makes a different type of RNA:

- RNA polymerase I: rRNA
- RNA polymerase II: mRNA
- RNA polymerase III: tRNA

The role of rRNA and tRNA will be covered later.

**RNA Polymerase II**

Like the prokaryotic polymerase, it contains multiple subunits.

- Two are similar to the $\beta$ and $\beta'$, and another two are like the $\alpha$
- In addition there are up to 12 additional subunits of varying size
- There is no equivalent of the $\sigma$ subunit
**Transcription factors**

In order to bind to DNA RNA pol II uses a series of proteins, termed transcription factors (TFIIA-D). These bind to the promoter and then allow the polymerase to bind.

The eukaryotic promoter sequences are similar to prokaryotic in outline, but differ in detail.

**Eukaryotic promoters**

There are typically two major sites:

-25 TATA box (similar to Pribnow box)
-80 CAAT box

Again, the exact sequence varies from gene to gene, but they determine where the polymerase will attach, and start to synthesize RNA.

**mRNA synthesis**

Once the polymerase has bound to a promoter, the actual synthesis of the mRNA is similar to that in prokaryotes, though there are some practical complications (DNA in chromosomes is harder to get at.) It is after the mRNA is made that major differences show up. Eukaryotic mRNA is not immediately ready for use. It must undergo extensive ‘post-transcriptional modification’ and then be transported out of the nucleus into the cytoplasm.
mRNA modifications

There are two distinct types of modification

Changes to the two ends of the mRNA (capping)

Removal of sections of RNA from the message (intron splicing)

When both of these are complete the mRNA is termed ‘mature’ and is ready to direct protein synthesis

5’ capping

The 5’ end of the mRNA is modified even before the rest of the molecule has been synthesized.

Function of 5’ cap

Three effects have been seen

Capping enhances the transport of mRNA out of the nucleus

mRNA degradation is slowed down because no nucleases recognize the 5-5 bond

Protein synthesis is enhanced (as much as 300x), though why is not clear
**Adding a polyA tail**

A string of about 200 adenines is added to the 3' end of the message.

It aids in protecting the message from degradation.

It attracts a polyA binding protein, which is important in efficient binding to ribosomes; so it increases the rate of protein synthesis.

**Introns**

One of the biggest surprises was the discovery of introns in 1977 by Phillip Sharp and Richard Roberts. For several years people had been puzzled by the presence of very large RNA molecules in the nucleus (hnRNA) that did not show up in the cytoplasm. Since these molecules had 3' and 5' caps, they looked like mRNA.

**mRNA/DNA hybrids**

Since mRNA molecules are copied from the DNA in a gene, they should have complementary sequences. As a result, if you mix mRNA with nuclear DNA, they should be able to form RNA:DNA hybrid molecules. But, when this was done, and the hybrids examined in an electron microscope, something odd was seen.
Hybridization problem

If the message is a direct copy of the gene, then the two should line up exactly.

3' Sense strand of original gene 5'

5' mRNA 3'

Hybridization problem

In fact the two only hybridized in a few areas.

There are large stretches of the gene that do not have a match in the mRNA.

The gene is about the same length as the hnRNAs.

What is going on?

The implication was clear: the gene is a lot bigger than the mature mRNA.

The original RNA is made as a copy of the entire gene: this is what the hnRNA molecules are.

The hnRNA molecules somehow become smaller, ending up as the mRNA seen in the cytoplasm.
From hnRNA to mRNA

Sharp and Roberts were able to show that the original RNA molecules had pieces cut out and discarded (introns). The remaining pieces (exons) were joined together (spliced) to form the mRNA seen in the cytoplasm.

1. Several snRNPs and proteins assemble to form a spliceosome. The 2' hydroxyl group on an adenine nucleotide (A) reacts with the 5' end of the intron, breaking RNA.

2. The 5' end of the intron becomes attached to the A nucleotide, forming a loop of RNA. The free 3' end of one exon reacts with the 5' end of the other.

3. The 3' and 5' ends of adjacent exons bond covalently, releasing the intron (which is then degraded).
What are introns?

The fact that most eukaryotic genes contain introns suggest they have an important role.

On the other hand, not all genes have them, and prokaryotes generally don’t have introns

Several suggestions have been made to explain their presence

Possible Role of introns

They are junk: they have no purpose but it is too dangerous to try and remove them

They are there to remove introns. This is true for the self-splicing ones, but is a circular argument. If they are not there, then you don’t need them

They are remnants left over from the production of new combinations of genetic information
**Functional domains**

Many proteins have one or more regions that have a specific purpose, such as Ca\(^{2+}\) binding. These are known as functional domains. If a new gene is created by bringing together two different functional domains from other genes, then it may pick up unwanted material at the same time.

**Intron formation**

More recently it appears that one advantage of introns is to allow the production of several different proteins from a single gene, by splicing the message in different ways. There seems to be evidence that this happens in many genes.

**Alternative splicing**
Using the mature message

Once the mRNA has been made, it is time for it to be used. Since most mRNAs are broken down quickly (in bacteria they have a half-life of just a few minutes) they need to be used in the process of protein synthesis as soon as possible.