DNA Transcription
By: Suzanne Clancy, Ph.D. © 2008 Nature Education
Citation: Clancy, S. (2008) DNA transcription. Nature Education 1(1)

If DNA is a book, then how is it read? Learn more about the DNA transcription process, where DNA is converted to RNA, a more portable set of instructions for the cell.

The genetic code is frequently referred to as a "blueprint" because it contains the instructions a cell requires in order to sustain itself. We now know that there is more to these instructions than simply the sequence of letters in the nucleotide code, however. For example, vast amounts of evidence demonstrate that this code is the basis for the production of various molecules, including RNA and protein. Research has also shown that the instructions stored within DNA are "read" in two steps: transcription and translation. In transcription, a portion of the double-stranded DNA template gives rise to a single-stranded RNA molecule. In some cases, the RNA molecule itself is a "finished product" that serves some important function within the cell. Often, however, transcription of an RNA molecule is followed by a translation step, which ultimately results in the production of a protein molecule.

Visualizing Transcription

The process of transcription can be visualized by electron microscopy (Figure 1); in fact, it was first observed using this method in 1970. In these early electron micrographs, the DNA molecules appear as "trunks," with many RNA "branches" extending out from them. When DNase and RNase (enzymes that degrade DNA and RNA, respectively) were added to the molecules, the application of DNase eliminated the trunk structures, while the use of RNase wiped out the branches.

DNA is double-stranded, but only one strand serves as a template for transcription at any given time; the other strand is referred to as the noncoding strand. In most organisms, the strand of DNA that serves as the coding template for one gene may be noncoding for other genes within the same chromosome.

The Transcription Process

The process of transcription begins when an enzyme called RNA polymerase (RNA pol) attaches to the template DNA strand and begins to catalyze production of complementary RNA. Polymerases are large enzymes composed of approximately a dozen subunits, and when active on DNA, they are also typically complexed with other factors. In many cases, these factors signal which gene is to be transcribed.

Three different types of RNA polymerase exist in eukaryotic cells, whereas bacteria have only one. In eukaryotes, RNA pol I transcribes the genes that encode most of the ribosomal RNAs (rRNAs), and RNA pol III transcribes the genes for one small rRNA, plus the transfer RNAs that play a key role in the translation process, as well as other small regulatory RNA molecules. Thus, it is RNA pol II that transcribes the messenger RNAs, which serve as the templates for production of protein molecules.

Transcription Initiation

The first step in transcription is initiation, when the RNA pol binds to the DNA upstream (5') of the gene at a specialized sequence called a promoter (Figure 2a). In bacteria, promoters are usually composed of three sequence elements, whereas in eukaryotes, there are as many as seven elements.
In prokaryotes, most genes have a sequence called the Pribnow box, with the consensus sequence TATAAT positioned about ten base pairs away from the site that serves as the location of transcription initiation. Not all Pribnow boxes have this exact nucleotide sequence; these nucleotides are simply the most common ones found at each site. Although substitutions do occur, each box nonetheless resembles this consensus fairly closely. Many genes also have the consensus sequence TTGCCA at a position 35 bases upstream of the start site, and some have what is called an upstream element, which is an A-T rich region 40 to 60 nucleotides upstream that enhances the rate of transcription (Figure 3). In any case, upon binding, the RNA pol "core enzyme" binds to another subunit called the sigma subunit to form a holozyme capable of unwinding the DNA double helix in order to facilitate access to the gene. The sigma subunit conveys promoter specificity to RNA polymerase; that is, it is responsible for telling RNA polymerase where to bind. There are a number of different sigma subunits that bind to different promoters and therefore assist in turning genes on and off as conditions change.

Eukaryotic promoters are more complex than their prokaryotic counterparts, in part because eukaryotes have the aforementioned three classes of RNA polymerase that transcribe different sets of genes. Many eukaryotic genes also possess enhancer sequences, which can be found at considerable distances from the genes they affect. Enhancer sequences control gene activation by binding with activator proteins and altering the 3-D structure of the DNA to help "attract" RNA pol II, thus regulating transcription. Because eukaryotic DNA is tightly packaged as chromatin, transcription also requires a number of specialized proteins that help make the coding strand accessible.

In eukaryotes, the "core" promoter for a gene transcribed by pol II is most often found immediately upstream (5') of the start site of the gene. Most pol II genes have a TATA box (consensus sequence TATTAA) 25 to 35 bases upstream of the initiation site, which affects the transcription rate and determines location of the start site. Eukaryotic RNA polymerases use a number of essential cofactors (collectively called general transcription factors), and one of these, TFIIID, recognizes the TATA box and ensures that the correct start site is used. Another cofactor, TFIIIB, recognizes a different common consensus sequence, G/C G/C G/C G/C G/C, approximately 38 to 32 bases upstream (Figure 4).

---

**Figure 4:** The promoters of genes transcribed by RNA polymerase II consist of a core promoter and a regulatory promoter that contain consensus sequences. Not all the consensus sequences shown are found in all promoters.
The terms "strong" and "weak" are often used to describe promoters and enhancers, according to their effects on transcription rates and thereby on gene expression. Alteration of promoter strength can have deleterious effects upon a cell, often resulting in disease. For example, some tumor-promoting viruses transform healthy cells by inserting strong promoters in the vicinity of growth-stimulating genes, while translocations in some cancer cells place genes that should be "turned off" in the proximity of strong promoters or enhancers.

Enhancer sequences do what their name suggests: They act to enhance the rate at which genes are transcribed, and their effects can be quite powerful. Enhancers can be thousands of nucleotides away from the promoters with which they interact, but they are brought into proximity by the looping of DNA. This looping is the result of interactions between the proteins bound to the enhancer and those bound to the promoter. The proteins that facilitate this looping are called activators, while those that inhibit it are called repressors.

Transcription of eukaryotic genes by polymerases I and III is initiated in a similar manner, but the promoter sequences and transcriptional activator proteins vary.

**Strand Elongation**

Once transcription is initiated, the DNA double helix unwinds and RNA polymerase reads the template strand, adding nucleotides to the 3' end of the growing chain (Figure 2b). At a temperature of 37 degrees Celsius, new nucleotides are added at the rate of about 15-20 amino acids per second in bacteria (Dennis & Bremer, 1974), while eukaryotes proceed at a much slower pace of approximately five to eight amino acids per second (Izban & Luse, 1992).

**Transcription Termination**

Terminator sequences are found close to the ends of coding sequences (Figure 2c). Bacteria possess two types of these sequences. In rho-independent terminators, inverted repeat sequences are transcribed; they can then fold back on themselves in hairpin loops, causing RNA pol to pause and resulting in release of the transcript. On the other hand, rho-dependent terminators make use of a factor called rho, which actively unwinds the DNA-RNA hybrid formed during transcription, thereby releasing the newly synthesized RNA (Figure 5).

In eukaryotes, termination of transcription occurs by different processes, depending upon the exact polymerase utilized. For pol I genes, transcription is stopped using a termination factor, through a mechanism similar to rho-dependent termination in bacteria. Transcription of pol III genes ends after transcribing a termination sequence that includes a polyuracil stretch, by a mechanism resembling rho-independent prokaryotic termination. Termination of pol II transcripts, however, is more complex.

Transcription of pol II genes can continue for hundreds or even thousands of nucleotides beyond the end of a coding sequence. The RNA strand is then cleaved by a complex that appears to associate with the polymerase. Cleavage seems to be coupled with termination of transcription and occurs at a consensus sequence. Mature pol II mRNAs are polyadenylated at the 3'-end, resulting in a poly(A) tail; this process follows cleavage and is also coordinated with termination.

Both polyadenylation and termination make use of the same consensus sequence, and the interdependence of the processes was demonstrated in the late 1980s by work from several groups. One group of scientists working with mouse globin genes showed that introducing mutations into the consensus sequence AATAAA, known to be necessary for poly(A) addition, inhibited both polyadenylation and transcription termination. They measured the extent of termination by hybridizing transcripts with the different poly(A) consensus sequence mutants with wild-type transcripts, and they were able to see a decrease in the signal of hybridization, suggesting that proper termination was inhibited. They therefore concluded that polyadenylation was necessary for termination (Logan et. al., 1987). Another group obtained similar results using a monkey viral
system, SV40 (simian virus 40). They introduced mutations into a poly(A) site, which caused mRNAs to accumulate to levels far above wild type (Connelly & Manley, 1988).

The exact relationship between cleavage and termination remains to be determined. One model supposes that cleavage itself triggers termination; another proposes that polymerase activity is affected when passing through the consensus sequence at the cleavage site, perhaps through changes in associated transcriptional activation factors. Thus, research in the area of prokaryotic and eukaryotic transcription is still focused on unraveling the molecular details of this complex process, data that will allow us to better understand how genes are transcribed and silenced.

Figure 5: Termination by bacterial rho-independent terminators is a multistep process.
Termination by bacterial rho-independent terminators is a multistep process

References and Recommended Reading


doi:10.1038/nature00769 (link to article)


