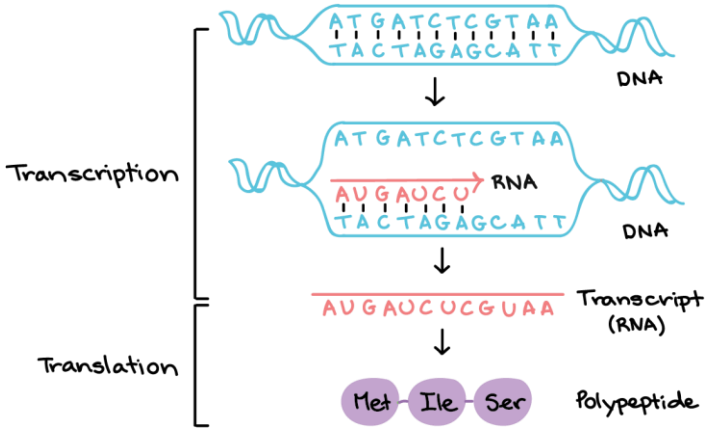
Overview of transcription

Transcription is the first step of gene expression. During this process, the DNA sequence of a gene is copied into RNA. Before transcription can take place, the DNA double helix must unwind near the gene that is getting transcribed. The region of opened-up DNA is called a **transcription bubble**.



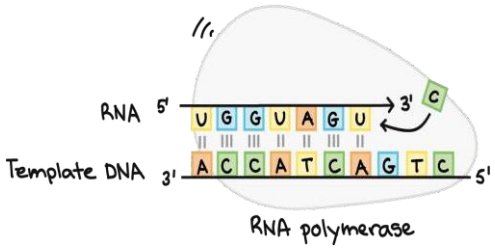
The goal of transcription is to make a RNA copy of a gene's DNA sequence. For a protein-coding gene, the RNA copy, or transcript, carries the information needed to build a polypeptide (protein or protein subunit). Eukaryotic transcripts need to go through some processing steps before translation into proteins.

Transcription uses one of the two exposed DNA strands as a template; this strand is called the template strand. The RNA product is complementary to the template strand and is almost identical to the other DNA strand, called the nontemplate (or coding) strand. However, there is one important difference: in the newly made RNA, all of the T nucleotides are replaced with U nucleotides.

The site on the DNA from which the first RNA nucleotide is transcribed is called the +1 site, or the initiation site. Nucleotides that come before the initiation site are given negative numbers and said to be upstream. Nucleotides that come after the initiation site are marked with positive numbers and said to be downstream. If the gene that's transcribed encodes a protein (which many genes do), the RNA molecule will be read to make a protein in a process called translation.

RNA polymerase

The main enzyme involved in transcription is **RNA polymerase**, which uses a single-stranded DNA template to synthesize a complementary strand of RNA. Specifically, RNA polymerase builds an RNA strand in the 5' to 3' direction, adding each new nucleotide to the 3' end of the strand.



RNA polymerases are enzymes that transcribe DNA into RNA. Using a DNA template, RNA polymerase builds a new RNA molecule through base pairing. For instance, if there is a G in the DNA template, RNA

polymerase will add a C to the new, growing RNA strand. RNA polymerase synthesizes an RNA strand complementary to a template DNA strand. It synthesizes the RNA strand in the 5' to 3' direction, while reading the template DNA strand in the 3' to 5' direction. The template DNA strand and RNA strand are antiparallel. RNA polymerases are large enzymes with multiple subunits, even in simple organisms like bacteria. In addition, humans and other eukaryotes have three different kinds of RNA polymerases: I, II, and III. Each one specializes in transcribing certain classes of genes.

Stages of transcription

Transcription of a gene takes place in **three stages: initiation, elongation, and termination**. Here, we will briefly see how these steps happen in bacteria. You can learn more about the details of each stage (and about how eukaryotic transcription is different) in the stages of transcription article.

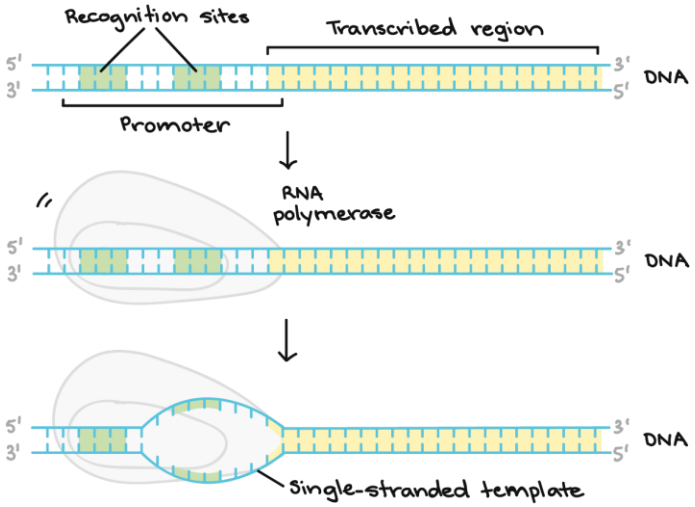
Initiation.

To begin transcribing a gene, RNA polymerase binds to the DNA of the gene at a region called the promoter. Basically, the promoter tells the polymerase where to "sit down" on the DNA and begin transcribing. RNA polymerase binds to a sequence of DNA called the promoter, found near the beginning of a gene. Each gene (or group of co-transcribed genes, in bacteria) has its own promoter. Once bound, RNA polymerase separates the DNA strands, providing the single-stranded template needed for transcription. Each gene (or, in bacteria, each group of genes transcribed together) has its own promoter. A promoter contains DNA sequences that let RNA polymerase or its helper proteins attach to the DNA. Once the transcription bubble has formed, the polymerase can start transcribing.

Promoters in bacteria

To get a better sense of how a promoter works, let's look an example from bacteria. A typical bacterial promoter contains two important DNA sequences, the -10 and -35 elements.

RNA polymerase recognizes and binds directly to these sequences. The sequences position the polymerase in the right spot to start transcribing a target gene, and they also make sure it's pointing in the right direction.

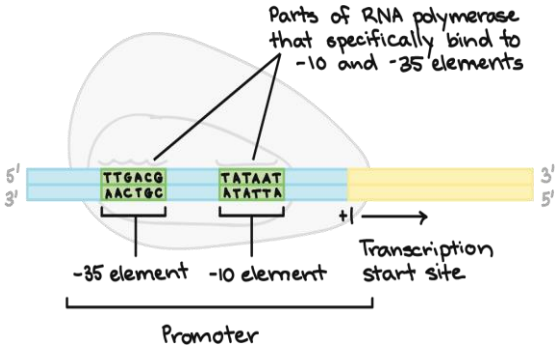


Once the RNA polymerase has bound, it can open up the DNA and get to work. DNA opening occurs at the -10 element, where the strands are easy to separate due to the many As and Ts (which bind to each other using just two hydrogen bonds, rather than the three hydrogen bonds of Gs and Cs).

The -10 and the -35 elements get their names because they come 35 and 10 nucleotides before the initiation site (+1in the DNA). The minus signs just mean that they are before, not after, the initiation site.

Promoters in humans

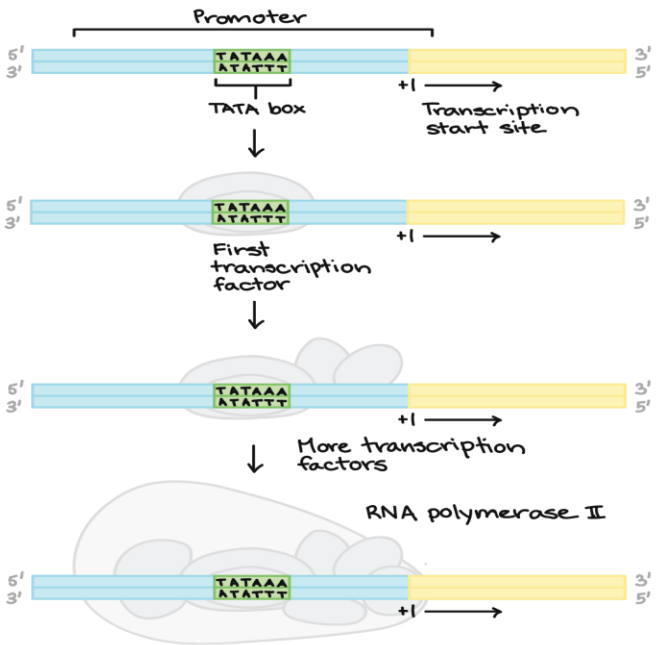
In eukaryotes like humans, the main RNA polymerase in your cells does not attach directly to promoters like bacterial RNA polymerase. Instead, helper proteins called basal (general) transcription factors bind to the promoter first, helping the RNA polymerase in your cells get a foothold on the DNA.



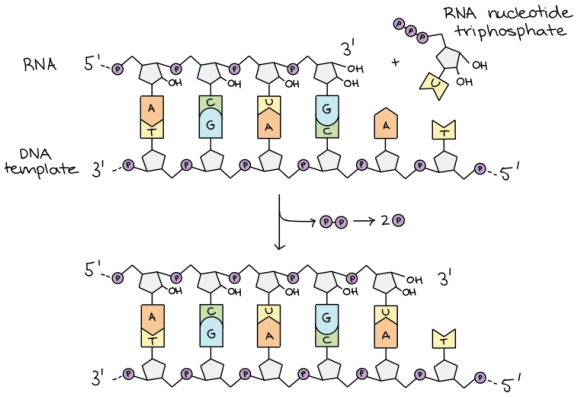
Many eukaryotic promoters have a sequence called a TATA box. The TATA box plays a role much like that of the -10 element in bacteria. It's recognized by one of the general transcription factors, allowing other transcription factors and eventually RNA polymerase to bind. It also contains lots of As and Ts, which make it easy to pull the strands of DNA apart.

Elongation.

One strand of DNA, the template strand, acts as a template for RNA polymerase. As it "reads" this template one base at a time, the polymerase builds an RNA molecule out of complementary nucleotides, making a chain that grows from 5' to 3'. The RNA transcript carries the same information as the non-template (coding) strand of DNA, but it contains the base uracil (U) instead of thymine (T). During

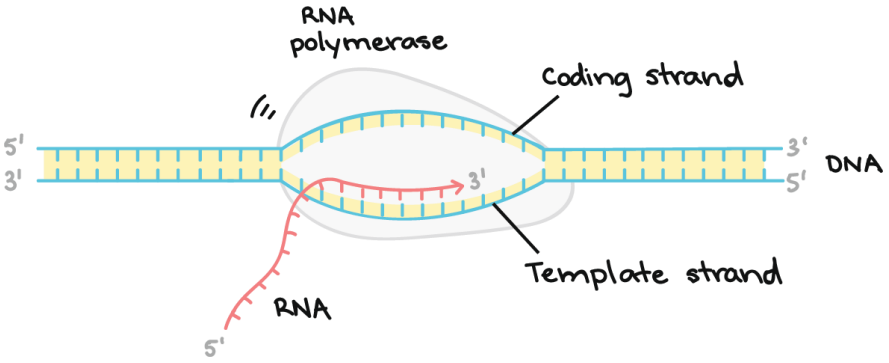


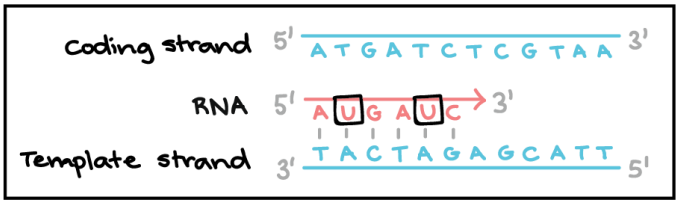
elongation, RNA polymerase "walks" along one strand of DNA, known as the template strand, in the 3' to 5' direction. For each nucleotide in the template, RNA polymerase adds a matching (complementary) RNA nucleotide to the 3' end of the RNA strand.

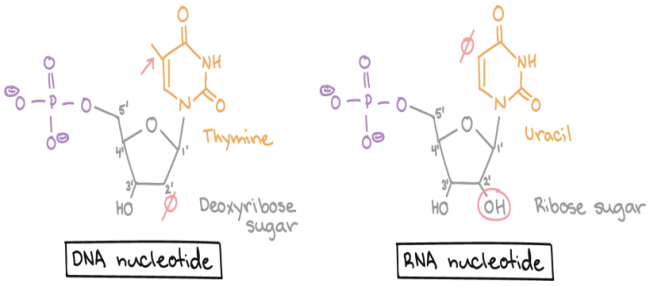


Here is the reaction that adds an RNA nucleotide to the chain:

The RNA transcript is nearly identical to the non-template, or coding, strand of DNA. However, RNA strands have the base uracil (U) in place of thymine (T), as well as a slightly different sugar in the nucleotide. So, as we can see in the diagram above, each T of the coding strand is replaced with a U in the RNA transcript. RNA nucleotides are similar to DNA nucleotides, but not identical. They have a ribose sugar rather than deoxyribose, so they have a hydroxyl group on the 2' carbon of the sugar ring. Also, in RNA, there is no T (thymine). Instead, RNA nucleotides carry the base uracil (U), which is structurally similar to thymine and forms complementary base pairs with adenine (A).





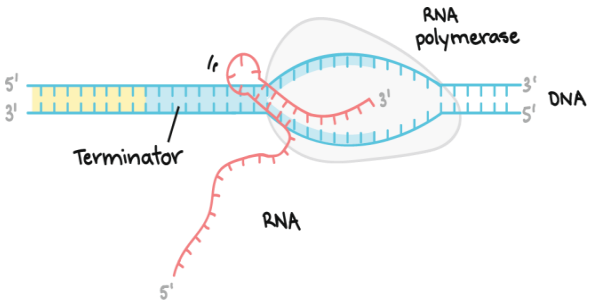


Termination.

Sequences called terminators signal that the RNA transcript is complete. Once they are transcribed, they cause the transcript to be released from the RNA polymerase. An example of a termination mechanism involving formation of a hairpin in the RNA is shown below.

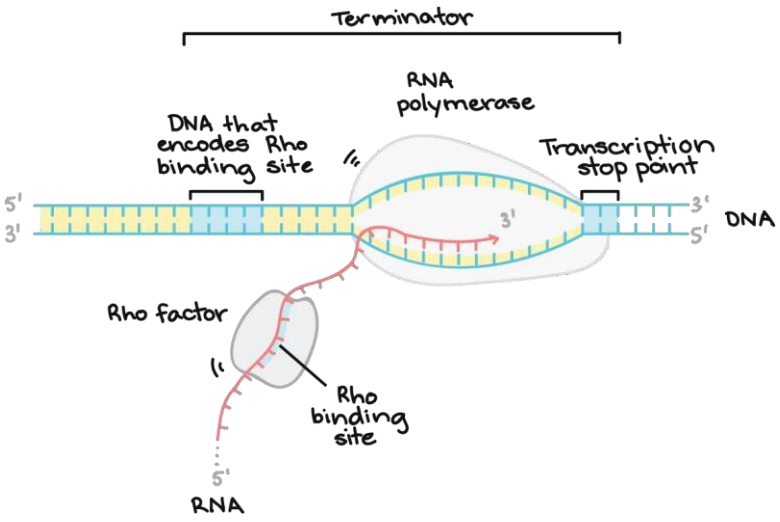
Termination in bacteria

There are two major termination strategies found in bacteria: Rho-dependent and Rho-independent. In Rho-dependent termination, the RNA contains a binding site for a protein called Rho factor. Rho factor binds to this sequence and starts "climbing" up the transcript towards RNA polymerase.

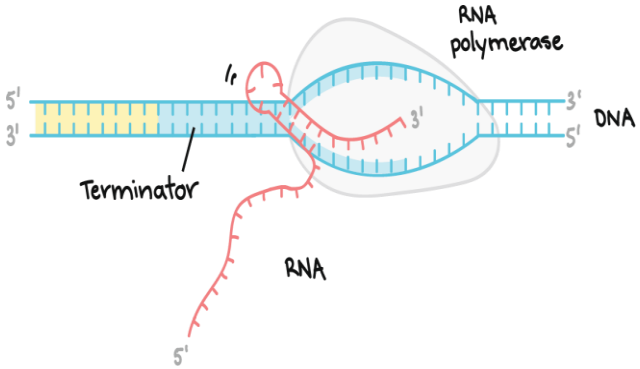


When it catches up with the polymerase at the transcription bubble, Rho pulls the RNA transcript and the template DNA strand apart, releasing the RNA molecule and ending transcription. Another sequence found later in the DNA, called the transcription stop point, causes RNA polymerase to pause and thus helps Rho catch up.

Rho-independent termination depends on specific sequences in the DNA template strand. As the RNA polymerase approaches the end of the gene being transcribed, it hits a region rich in C and G nucleotides. The RNA transcribed from this region folds back on itself, and the complementary C and G nucleotides bind together. The result is a stable hairpin that causes the polymerase to stall.



In a terminator, the hairpin is followed by a stretch of U nucleotides in the RNA, which match up with A nucleotides in the template DNA. The complementary U-A region of the RNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, produces enough instability for the enzyme to fall off and liberate the new RNA transcript.

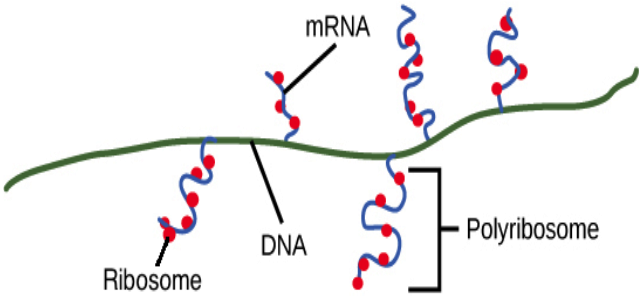


In eukaryotes like humans, transcription termination happens differently depending on the type of gene involved. Here, we'll see how termination works for protein-coding genes. Termination begins when a polyadenylation signal appears in the RNA transcript. This is

a sequence of nucleotides that marks where an RNA transcript should end. The polyadenylation signal is recognized by an enzyme that cuts the RNA transcript nearby, releasing it from RNA polymerase. Oddly enough, RNA polymerase continues transcribing after the transcript is released, often making 500 - 2 000 more nucleotides' worth of RNA. Eventually, it detaches from the DNA through mechanisms that are not yet fully understood. The extra RNA is not usually translated and seems to be a wasteful byproduct of transcription.

What happens to the RNA transcript?

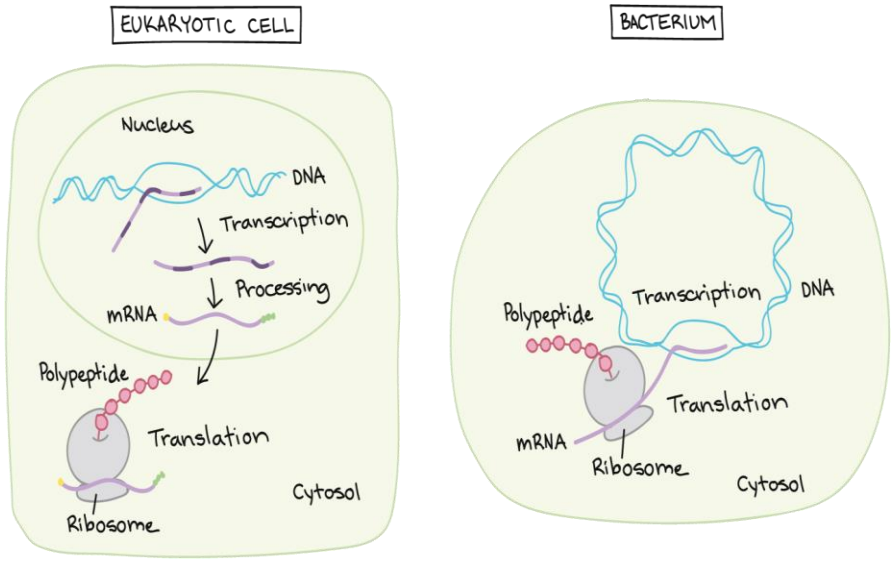
After termination, transcription is finished. An RNA transcript that is ready to be used in translation is called a messenger RNA (mRNA). In bacteria, RNA transcripts are ready to be translated right after transcription. In fact, they're actually ready a little sooner than that: translation may start while transcription is still going on!



In the diagram below, mRNAs are being transcribed from several different genes. Although transcription is still in progress, ribosomes have attached each mRNA and begun to translate it into protein. When an mRNA is being translated by multiple ribosomes, the mRNA and ribosomes together are said to form a polyribosome.

Why can transcription and translation happen simultaneously for an mRNA in bacteria?

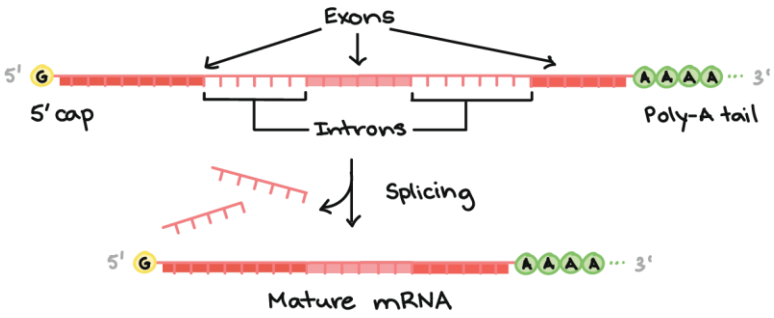
One reason is that these processes occur in the same 5' to 3' direction. That means one can follow or "chase" another that's still occurring. Also, in bacteria, there are no internal membrane compartments to separate transcription from translation.



Transcription happens in the nucleus of human cells, while translation happens in the cytosol. Also, in eukaryotes, RNA molecules need to go through special processing steps before translation. That means translation can't start until transcription and RNA processing are fully finished.

Overview of pre-mRNA processing in eukaryotes

As a quick review, gene expression (the "reading out" of a gene to make a protein, or chunk of a protein) happens a little bit differently in bacteria and eukaryotes such as humans.



In bacteria, RNA transcripts are ready to act as messenger RNAs and get translated into proteins right away. In eukaryotes, things are a little more complex, though in an pretty interesting way. The molecule that's directly made by transcription in one of your (eukaryotic) cells is called a pre-mRNA, reflecting that it needs to go through a few more steps to become an actual messenger RNA (mRNA). These are: Addition of a 5' cap to the beginning of the RNA Addition of a poly-A tail (tail of A nucleotides) to the end of the RNA Chopping out of introns, or "junk" sequences, and pasting together of the remaining, good sequences (exons) Once it's completed these steps, the RNA is a mature mRNA. It can travel out of the nucleus and be used to make a protein.

5' cap and poly-A tail

Both ends of a pre-mRNA are modified by the addition of chemical groups. The group at the beginning (5' end) is called a cap, while the group at the end (3' end) is called a tail. Both the cap and the tail protect the transcript and help it get exported from the nucleus and translated on the ribosomes (protein-making "machines") found in the cytosol. The 5’ cap is added to the first nucleotide in the transcript during transcription. The cap is a modified guanine (G) nucleotide, and it protects the transcript from being broken down. It also helps the ribosome attach to the mRNA and start reading it to make a protein.

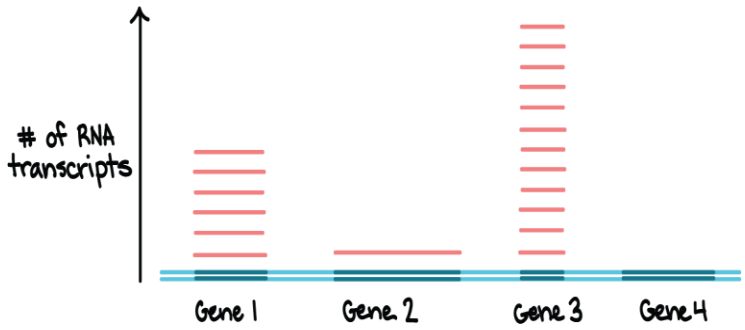
Eukaryotic RNA modifications

In bacteria, RNA transcripts can act as messenger RNAs (mRNAs) right away. In eukaryotes, the transcript of a protein-coding gene is called a pre-mRNA and must go through extra processing before it can direct translation. Eukaryotic pre-mRNAs must have their ends modified, by addition of a 5' cap (at the beginning) and 3' poly-A tail (at the end). Many eukaryotic pre-mRNAs undergo splicing. In this process, parts of the pre-mRNA (called introns) are chopped out, and the remaining pieces (called exons) are stuck back together. End modifications increase the stability of the mRNA, while splicing gives the mRNA its correct sequence. (If the introns are not removed, they'll be translated along with the exons, producing a "gibberish" polypeptide.)

To learn more about pre-mRNA modifications in eukaryotes, check out the article on pre-mRNA processing.

Transcription happens for individual genes

Not all genes are transcribed all the time. Instead, transcription is controlled individually for each gene (or, in bacteria, for small groups of genes that are transcribed together). Cells carefully regulate transcription, transcribing just the genes whose products are needed at a particular moment.



For example, the diagram below shows a "snapshot" of an imaginary cell's RNAs at a given moment in time. In this cell, genes 1, 2 and 3, are transcribed, while gene 4 is not. Also, genes 1, 2, and 3 are transcribed at different levels, meaning that different numbers of RNA molecules are made for each.

In the following articles, we'll take a more in-depth look at RNA polymerase, the stages of transcription, and the process of RNA modification in eukaryotes. We'll also consider some important differences between bacterial and eukaryotic transcription.

Stages of transcription

RNA polymerase is crucial because it carries out transcription, the process of copying DNA (deoxyribonucleic acid, the genetic material) into RNA (ribonucleic acid, a similar but more short-lived molecule). Transcription is an essential step in using the information from genes in our DNA to make proteins. Proteins are the key molecules that give cells structure and keep them running. Blocking transcription with mushroom toxin causes liver failure and death, because no new RNAs—and thus, no new proteins—can be made. Transcription is essential to life, and understanding how it works is important to human health. Let's take a closer look at what happens during transcription.