Biochemistry 401 Lecture 37. Today we're
going to talk about transcription, both in bacterial and eukaryotic systems.

Hi, today we're going to talk about transcription. This is the synthesis of RNA, through the polymerization of ribonucleoside triphosphates by using DNA as a template.

Before we get started, it's important to get some terminology straight, so that we all start on the same page. RNA is transcribed from genes.

When the genetic sequence is written with two strands, the top strand will be the coding strand and the bottom strand is the template strand. The top strand is called the coding strand,

because it's the same sequence as the primary RNA transcript is going to be, except that the RNA is going to include U's instead of T's. The bottom strand is the strand that's going to be used for the template for the synthesis of RNA.

We also have to have some terminology so that we can talk about directions in relation to a landmark. In transcription, there directions are upstream, which is in the five prime direction and downstream, which is in the three prime direction, and this is in relation to a landmark, and it's also in relation to the top strand.

Why? Because in general, the top strand the coding strand is the more important sequence. For instance, when we write a genetic sequence and only use one strand, we write the coding strand, 5 prime to 3 prime. We don't write the template strand, generally speaking, and so when we're talking about terminology that includes upstream and downstream, we're talking about the five prime direction, and the three prime direction, respectively, in relation to the coding strand – the top strand. I hope that's clear.

There are three main stages of RNA synthesis – initiation elongation and termination. Initiation is beginning, and in order to begin transcription, the polymerase, the enzyme that's going to synthesize this RNA, has to find out where to start. So it has to find a region in the DNA that's called the promoter. It's going to bind to the promoter in order to initiate transcription. So, promoter recognition, and then we have to
form an open promoter that is melting the DNA strands so that initiation can
proceed. in order to begin transcription, the polymerase actually has to leave the
promoter so that's called promoter escape. And then we have elongation, and
finally we have termination, when RNA synthesis stops. And so that's initiation,
elongation, and termination.

We're going to look at the synthesis of RNA in prokaryotes like the bacteria *E.coli*, and also in eukaryotes, and we're going to use humans as a model system for that.

There are similarities and there are differences in these two types of transcription those in prokaryotes like bacteria and those in eukaryotes, like humans. The similarities for both systems are shown here. Both required template DNA that have promoter and termination signals. they both require Ribonucleoside triphosphates, and these are AGC and U. We have U's instead of T's, remember. And polymerases in both systems require magnesium as a cofactor. And finally, both systems require a polymerase but what isn't needed is a primer. Primers are not needed.

Since we're using RNA for transcription anyway, we don't need an RNA primer. The actual process of polymerizing this RNA is also similar in prokaryotes and eukaryotes, in this respect. The template strand is going to be read 3 prime to 5 prime and the RNA itself is going to be synthesized 5 prime to 3 prime in a fashion that's complementary to the template. The synthesis of this RNA is going to be driven forward by loss of pyrophosphate and subsequent hydrolysis of the pyrophosphate to 2 molecules of inorganic phosphate is the same in prokaryotes and in eukaryotes. And another thing is, is that we need magnesium in order to position these substrates at the right place inside the polymerase and also to help catalyze the reaction itself. So magnesium is required as a cofactor.

But there are many differences, too. The polymerases that are used in prokaryotes and in eukaryotes are different. For instance, there's one main polymerase for transcription in bacteria, but there are three main polymerases that are used in eukaryotes – Pol I, Pol II, and Pol III. We'll talk about those later. So the polymerases are different. The way that RNA synthesis is initiated is also different. It's much more complex in eukaryotes than it is in bacteria, and termination is also different. Eukaryotes also use a lot of post-transcriptional modifications in their RNA, especially messenger RNA. That's the RNA that we use to make proteins. That's substantially modified. We'll look at that in a minute, too. And another thing that's different, is that transcription happens in the same place as translation in prokaryotes, like bacteria and archaea. That's because the genomic DNA is situated in the cytosol. There's no intervening nuclear envelope. And so, the transcription and translation can happen at the same time on the same transcript. As a transcript's being made, it can also be translated. That's different in eukaryotes.

In eukaryotes, our genomic DNA is in the nucleus, and so we can synthesize messenger RNA for instance, but then that transcript has to be exported out of the nucleus, and into the cytosol for translation.

So let's talk about transcription in E. coli. This is a prokaryotic model system. And let's look at the polymerase first. This is the enzyme that's going to actually synthesize the RNA. This enzyme contains two major parts. It contains a targeting protein that is called the sigma factor. This is going to target the holoenzyme to the promoter. And it also contains the core polymerase. This is the portion of the enzyme that's actually going to catalyze the polymerization itself. Since the sigma factor is the "targeter", this is going to provide the specificity of binding for the polymerase to actually

find the right promoter, to transcribe the right gene at the right time. And so E.coli has one main RNA polymerase, but it contains several sigma factors, and it's the sigma factor that provides the specificity for transcription for bacteria. The core polymerase is made up of five main regions – alpha 1, alpha 2, beta, beta prime, and omega. We're not quite sure what omega does, but alpha 1 and alpha 2 are structural. Beta and beta prime are catalytic. Those subunits are going to be the ones that are actually going to catalyze the polymerization itself. These two round objects that we see on the alpha subunits are the CTDs – the c-terminal domains. These are very important for stabilizing the polymerase on the target DNA. We're going to look at that in just a minute.
So how does sigma know where the beginning of a gene is? How does it know where to go?
Well it goes to a region of the gene called the promoter region, and in bacteria the promoter region has two main parts. It has a minus 10 and a minus 35 region. What does that mean? Let's talk about that for a minute.

We have the start site of transcription – let's call that one. That is 1. And then when we're going upstream we have negative 1, negative 2, negative 3, negative 4, all the way up negative 10, and negative 35. But please note that there's no zero. There's one and there's four. Upstream is going to be minus one, minus two, minus three, The regions of the promoter that we're interested in for sigma binding are called the minus ten region and the minus 35 region. And the intervening space is also important between them. Different promoters have different lengths of DNA between these two regions. As you can see from this diagram, the minus 10 region is also called the Pribnow box. And this is what we call the minus 10 region, in E. coli. Both the minus 35 and the minus 10 regions are very important for attracting the sigma factor.

That's because the sigma factor binds specifically to these places on the promoter. The intervening space is also important between these two regions, because it will dictate whether that's a comfortable fit for the sigma factor or whether it's a strained fit.

If you notice at the bottom of the slide, there's something that's called the consensus sequence, and this was
determined by looking at the promoters of several bacterial genes, and saying what nucleotides are the same and what are different. And those nucleotides that were most often seen were noted, and the optimal space between these regions, and how far from the initiation site these two regions are, were also noted. And so this yielded what's called a consensus sequence – the most frequently seen nucleotides, the most frequently seen spacing – and promoters that are closer to this consensus sequence are stronger promoters than those that have deviations from this consensus sequence. Deviations from consensus yield a weaker