

# Information Processing: Transcription

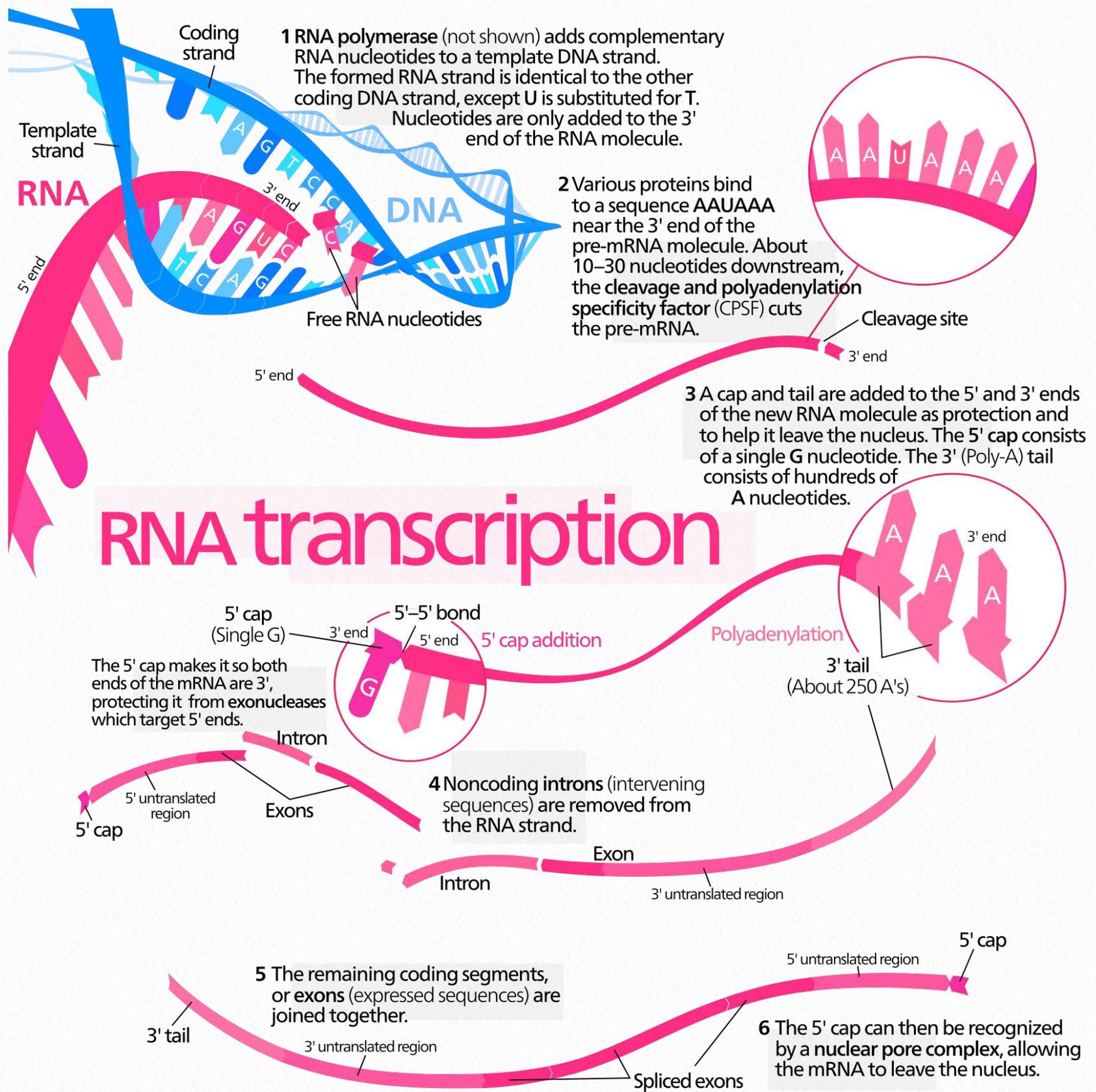


In the preceding sections, we have discussed the replication of the cell's DNA and the mechanisms by which the integrity of the genetic information is carefully maintained. What do cells do with this information? How does the sequence in DNA control what happens in a cell? If DNA is a giant instruction book containing all of the cell's "knowledge" that is copied and passed down from generation to generation, what are the instructions for? And how do cells use these instructions to make what they need?

## Information flow

You have learned in introductory biology courses that genes, which are instructions for making proteins, are made of DNA. You also know that information in genes is copied into temporary instructions called messenger RNAs that direct the synthesis of specific proteins. This description of flow of information from DNA to RNA to protein is often called the central dogma of molecular biology and is a good starting point for

**YouTube Lectures  
by Kevin  
[HERE](#) & [HERE](#)**



**Figure 7.52 - Overview of eukaryotic transcription**

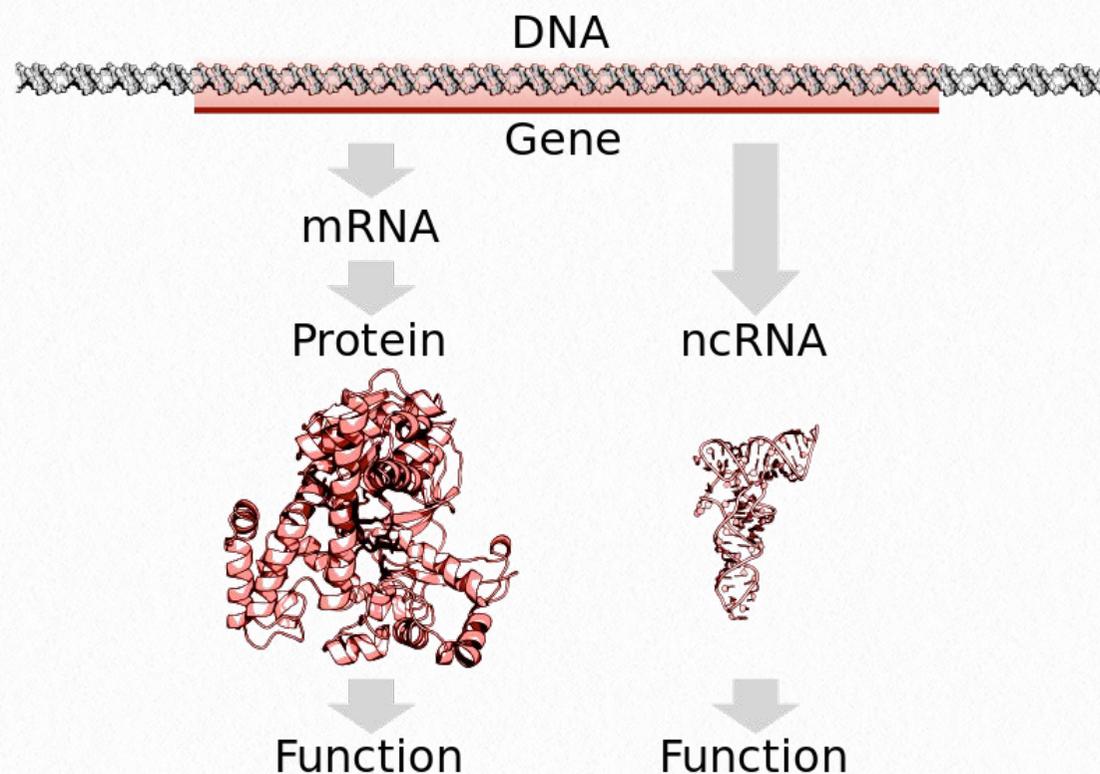
Wikipedia

an examination of how cells use the information in DNA.

Consider that all of the cells in a multicellular organism have arisen by division from a sin-

gle fertilized egg and therefore, all have the same DNA. Division of that original fertilized egg produces, in the case of humans, over a trillion cells, by the time a baby is produced from that egg (that's a lot of DNA replica-

tion!). Yet, we also know that a baby is not a giant ball of a trillion identical cells, but has the many different kinds of cells that make up tissues like skin and muscle and bone and nerves. How did cells that have identical DNA turn out so different?



**Figure 7.53 - Transcripts may code for protein or may be functional as RNAs**

Wikipedia

But, apart from copying one, rather than both strands of DNA, how is transcription different from replication of DNA? DNA replication serves to copy all the genetic material of the

The answer lies in gene expression, which is the process by which the information in DNA is used. Although all the cells in a baby have the same DNA, each different cell type uses a different subset of the genes in that DNA to direct the synthesis of a distinctive set of RNAs and proteins. The first step in gene expression is transcription, which we will examine next (Figure 7.52).

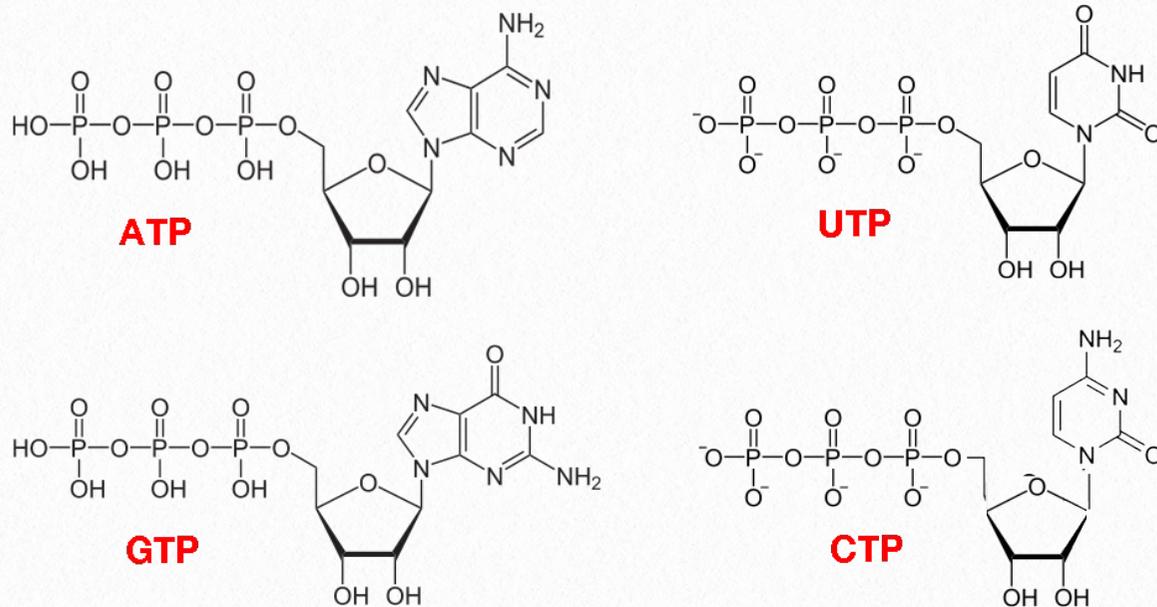
## Transcription

Transcription is the process of copying information from DNA sequences into RNA sequences. This process is also known as DNA-dependent RNA synthesis. When a sequence of DNA is transcribed, only one of the two DNA strands is copied into RNA. We will consider what determines which strand of DNA is copied into RNA, later on.

cell and occurs before a cell divides, so that a full copy of the cell's genetic information can be passed on to the daughter cell. Transcription, by contrast, copies short stretches of the coding regions of DNA to make RNA. Different genes may be copied into RNA at different times in the cell's life cycle. RNAs are, essentially, temporary copies of the information in DNA and different sets of instructions are copied for use at different times and in different cell types.

Cells make several different kinds of RNA:

- mRNAs that code for proteins
- rRNAs that form part of ribosomes



**Figure 7.54 - The four ribonucleotides for making RNA**

- tRNAs that serve as adaptors between mRNA and amino acids during translation
- Small RNAs that regulate gene expression, including miRNAs and siRNAs
- Other small RNAs that have a variety of functions, including the small nuclear RNAs that are part of the splicing machinery.
- Long noncoding RNAs (lnc RNAs - [Figure 7.53](#))

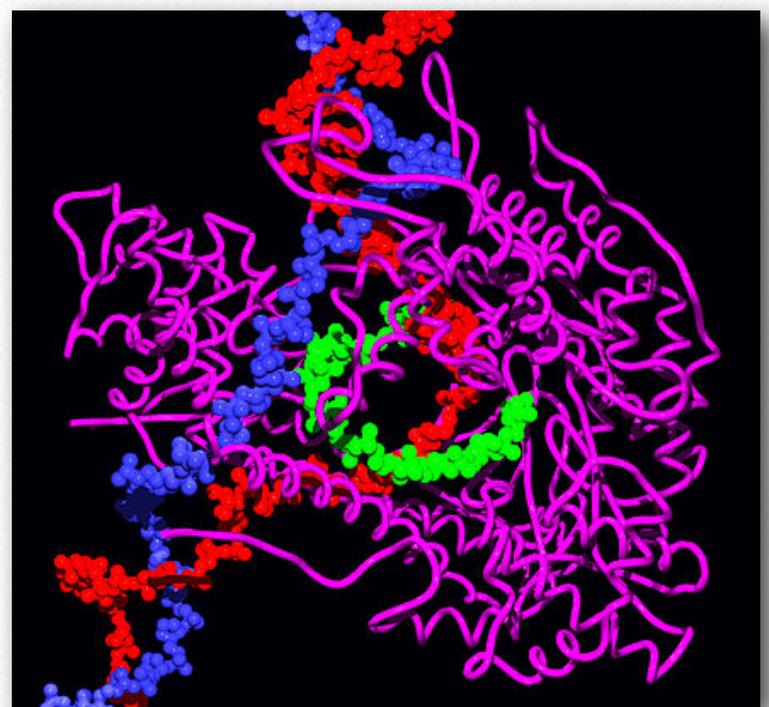
Building an RNA strand is very similar to building a DNA strand. This is not surprising, knowing that DNA and RNA are very similar molecules. Transcription is catalyzed by the enzyme RNA Polymerase. "RNA polymerase" is a general term for an enzyme that makes RNA. There are several different kinds of RNA polymerases in eukaryotic cells, while in prokaryotes, a single type of

RNA polymerase is responsible for all transcription.

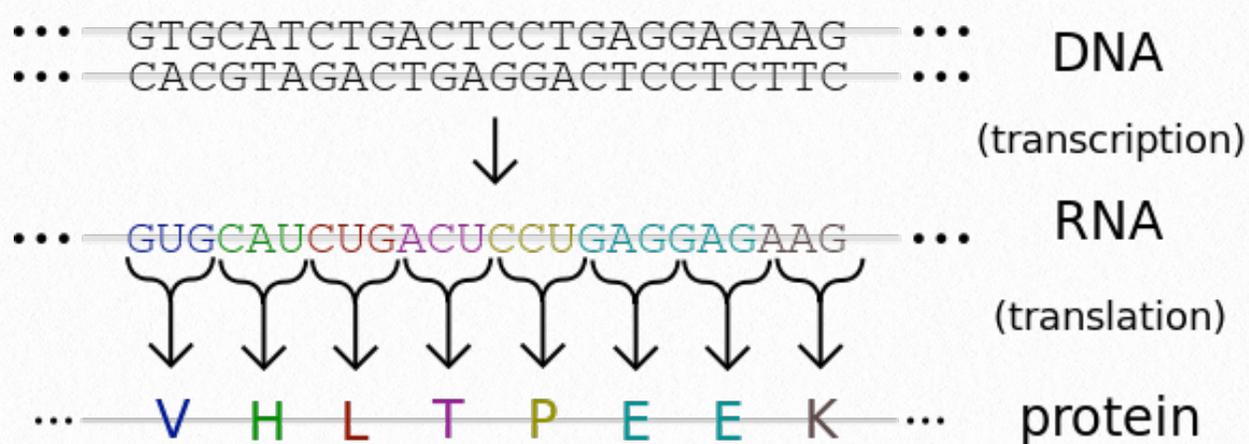
## RNA synthesis

Like DNA polymerases, RNA polymerases synthesize new strands only in the 5' to 3' direction, but because they are making RNA, they use ribonucleotides (i.e., RNA nucleotides - [Figure 7.54](#)) rather than deoxyribonu-

cleotides. Ribonucleotides are joined in exactly the same way as deoxyribonucleotides, i.e., the 3'OH of the last nucleotide on the growing chain is joined to the 5' phosphate on



**Figure 7.55 - RNA (green) being synthesized from DNA template (blue strand) by T7 RNA polymerase (purple). The non-template DNA strand is in red.**



**Figure 7.56 - Central dogma - DNA to RNA to protein**

Wikipedia

DNA. Much of this is non-coding DNA, meaning that it will not need to be transcribed. The small percentage of the genome that is made up of coding sequences still amounts to between 20,000 and 30,000 genes in each cell. Of these genes, only a small

the incoming nucleotide to make a phosphodiester bond.

One important difference between DNA polymerases and RNA polymerases is that the latter do not require a primer to start making RNA. Once RNA polymerases are in the right place to start copying DNA, they just begin making RNA by joining together RNA nucleotides complementary to the DNA template.

### Starting points

This, of course, brings us to an obvious question- how do RNA polymerases "know" where to start copying on the DNA?

Unlike the situation in replication, where every nucleotide of the parental DNA must eventually be copied, transcription, as we have already noted, only copies selected portions of the DNA into RNA at any given time. Consider the challenge here: in a human cell, there are approximately 6 billion base-pairs of

number will need to be expressed at any given time. What indicates to an RNA polymerase where to start copying DNA to make a transcript?

### Promoters

It turns out that patterns in the DNA sequence indicate where RNA polymerase should start and end transcription. These sequences are recognized by the RNA polymerase or by proteins that help RNA polymerase determine where it should bind the DNA to start transcription. A DNA sequence at which the RNA polymerase binds to start transcription is called a promoter. The DNA sequence that indicates the endpoint of transcription, where the RNA polymerase should stop adding nucleotides and dissociate from the template is known as a terminator sequence. The promoter and terminator, thus, bracket the region of the DNA that is to be transcribed.

Gene	-35 sequence	-10 sequence	Transcription start site
			+1
<i>rrnE1</i>	CAATTTTTC	TATTGAGGAATG	AGGAGAACTCCCTATAATGCGCCTCCATC
<i>tRNA<sup>tyr</sup></i>	CAACGTAAC	ACTTTACAGCGG	CGCGTCATTTGATATGATGCGCCCCGCT
<i>trp</i>	AAATGAGCTGTTGACA	AATTAATCATCGAACTAGTTAACTAGTACGCAAG	
<i>araBAD</i>	GGATCCT	ACCTGACGCTTTT	TATCGCAACTCTCTACTGTTTCTCCATAC
<i>araC</i>	GCCGTGATTATAGACACTTTT	TGTTACGCGTTTTT	TGTCATGGCTTTTGGTTC
<i>bioA</i>	TTCCAAAACGTGTTTTTTGTTG	TTAATTCGGTGTAGACTTGTAA	ACCTA

**Figure 7.57 - Sequences upstream of transcription start site in several prokaryotic genes**

Image by Martha Baker

A promoter is described as being situated upstream of the gene that it controls (Figure 7.57). What this means is that on the DNA strand that the gene is on, the promoter sequence is "before" the gene, or to put it differently, it is on the side of the gene opposite to the direction of transcription. Also notice that the promoter is said to "control" the gene it is associated with. This is because expression of the gene is dependent on the binding of RNA polymerase to the promoter sequence to begin transcription. If the RNA polymerase and its helper proteins do not bind at the promoter, the gene cannot be transcribed and it will therefore, not be expressed.

What is special about a promoter sequence? In an effort to answer this question, scientists examined many genes and their surrounding sequences (Figure 7.57). Because the same RNA polymerase has to bind to many differ-

ent promoters, it would be predicted that promoters would have some similarities in their sequences. As expected, common sequence patterns were seen to be present in many promoters.

We will first take a look at prokaryotic promoters.

When prokaryotic genes were examined, the following features commonly emerged:

- A transcription start site (this the base in the DNA across from which the first RNA nucleotide is paired), which, by convention, is denoted as +1.
- A -10 sequence: this is a 6 bp region centered about 10 bp upstream of the start site. The consensus sequence at this position is TATAAT. In other words, if you count back from the transcription start site, the se-

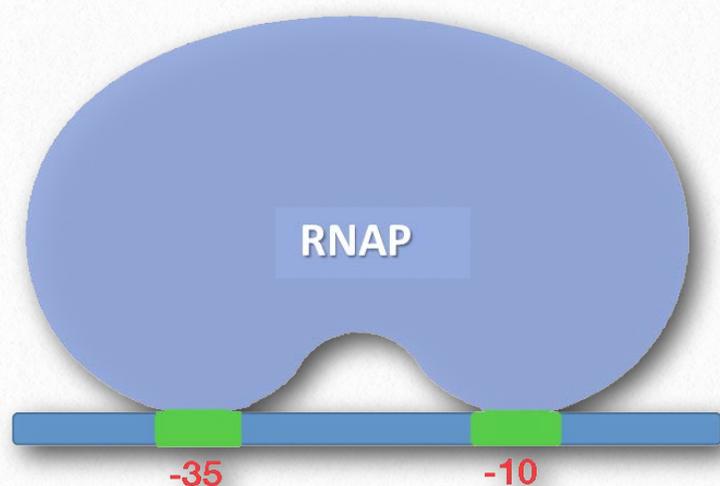


quence found at roughly -10 in the majority of promoters studied is TATAAT.

- A -35 sequence: this is a 6 bp sequence at about 35 basepairs upstream from the start of transcription. The consensus sequence at this position is TTGACA.

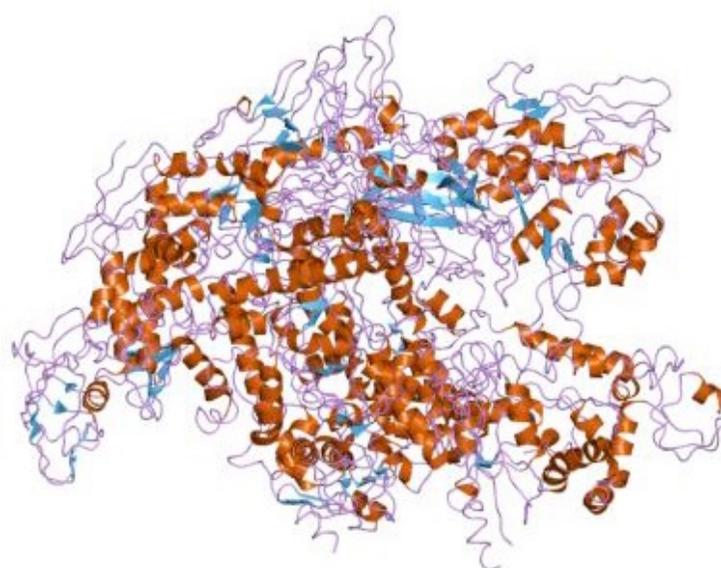
-10 Sequence					
T	A	T	A	A	T
77%	76%	60%	61%	56%	82%
-35 Sequence					
T	T	G	A	C	A
69%	70%	61%	56%	54%	54%

It is important to understand that each nucleotide in a consensus sequence is simply the one that appeared at that position in the majority of promoters examined, and does not mean that the entire consensus sequence is found in all promoters. In fact, few promoters have -10 and -35 sequences that exactly match the con-



**Figure 7.58 - RNA polymerase promoter binding**

Wikipedia



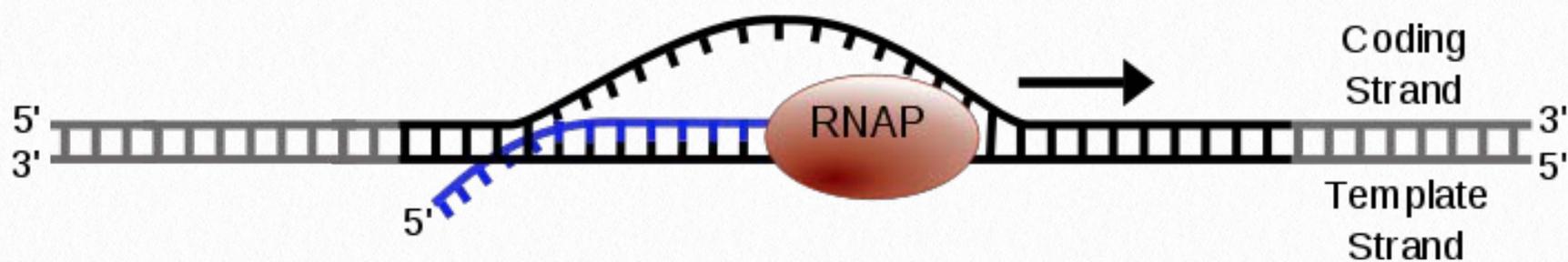
**Figure 7.59 - A bacterial RNA polymerase ( $\alpha 2\beta\beta'$  and  $\omega$ )**

sensus. The box at the left shows the -10 and -35 sequences by percentage of occurrence of each base in the promoter.

What is the significance of these sequences? It turns out that the sequences at -10 and -35 are necessary for recognition of the promoter region by RNA polymerase (Figure 7.58). The sequences at -10 and -35 may vary a little in individual promoters, as mentioned above, but the extent to which they are different is limited. It is only when the RNA polymerase has stably bound at the promoter that transcription can begin. The process by which the promoter is recognized and bound stably has been well studied for the RNA polymerase of *E. coli*.

### Core polymerase and holoenzyme

The *E. coli* RNA polymerase is made up of a core enzyme of five subunits ( $\alpha 2\beta\beta'$  and  $\omega$ ) and an additional subunit called the  $\sigma$



**Figure 7.60 - Synthesis of RNA in the transcription bubble**

(sigma) subunit. Together, the  $\sigma$  subunit and core polymerase make up what is termed the RNA polymerase holoenzyme. The core polymerase is the part of the RNA polymerase that is responsible for the actual synthesis of the RNA, while the  $\sigma$  subunit is necessary for binding of the enzyme at promoters to initiate transcription.

### Loose association

The core polymerase and  $\sigma$  subunit are not always associated with each other. For the most part, the core polymerase is loosely associated with DNA, although it does not discriminate between promoters and other sequences in DNA, and the DNA strands are not opened up to allow transcription in this state. The role of the  $\sigma$  subunit is to reduce the affinity of the core polymerase for non-specific DNA sequences and to help the enzyme specifically bind to promoter sequences.

### Holoenzyme binding

It is when the  $\sigma$  subunit associates with the core polymerase that the holoenzyme is able to bind specifically to promoter sequences. The initial binding of the holoen-

zyme at the promoter results in what is called a “closed” complex, meaning that the DNA template is still double-stranded and has not opened up to allow transcription. This closed complex is then converted to an “open” complex by the separation of the DNA strands to create a transcription bubble about 12-14 base-pairs long (Figure 7.60). The conversion of the closed complex to the open complex also requires the presence of the  $\sigma$  subunit.

### Open complex & initiation

Once the open complex has formed, the DNA template can begin to be copied, and the core polymerase adds nucleotides complementary to one strand of the DNA. At this stage, known as initiation, the polymerase adds several nucleotides while still bound to the promoter, and without moving along the DNA template. Initially, short pieces of RNA a few nucleotides long may be made and released, without the polymerase leaving the promoter. Eventually, the enzyme makes the transition to the next stage, elongation, when

an RNA of 8-9 bases is made and the enzyme moves beyond the promoter region.

## Elongation

Once elongation commences and the RNA polymerase is moving down the DNA template, the  $\sigma$  subunit is no longer necessary and may dissociate from the core enzyme. The core polymerase can move along the template, unwinding the DNA ahead of it to maintain a transcription bubble of 12-15 base-pairs and synthesizing RNA complementary to one of the strands of the DNA. As already mentioned, an RNA chain, complementary to the DNA template, is built by the RNA polymerase by the joining of the 5' phosphate of an incoming ribonucleotide to the 3'OH on the last nucleotide of the growing RNA strand. Behind the RNA polymerase, the DNA template is rewound, displacing the newly made RNA from its template strand.

## Termination

As mentioned earlier, a sequence of nucleotides called the terminator is the signal to

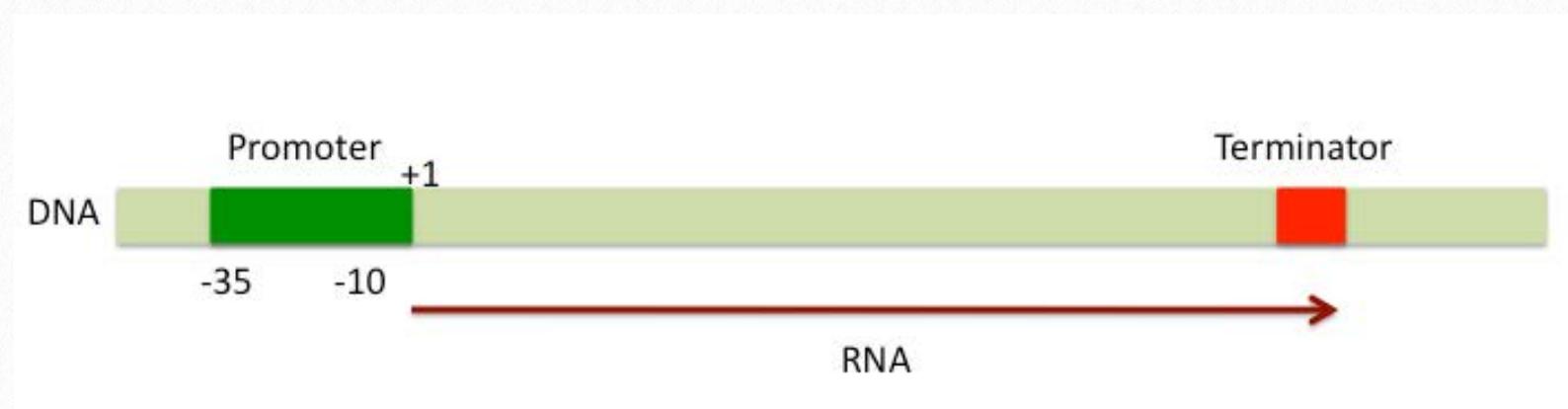
the RNA polymerase to stop transcription and dissociate from the template. Some terminator sequences, known as intrinsic terminators, allow termination by RNA polymerase without the help of any additional factors, while others, called rho-dependent terminators, require the assistance of a protein factor called rho ( $\rho$ ).

How does the sequence of the terminator cause the RNA polymerase to stop adding nucleotides and release the transcript?

To understand this, it is useful to know that the terminator sequence precedes the last nucleotide of the transcript. In other words, the terminator is part of the end of the sequence that is transcribed ([Figure 7.61](#)).

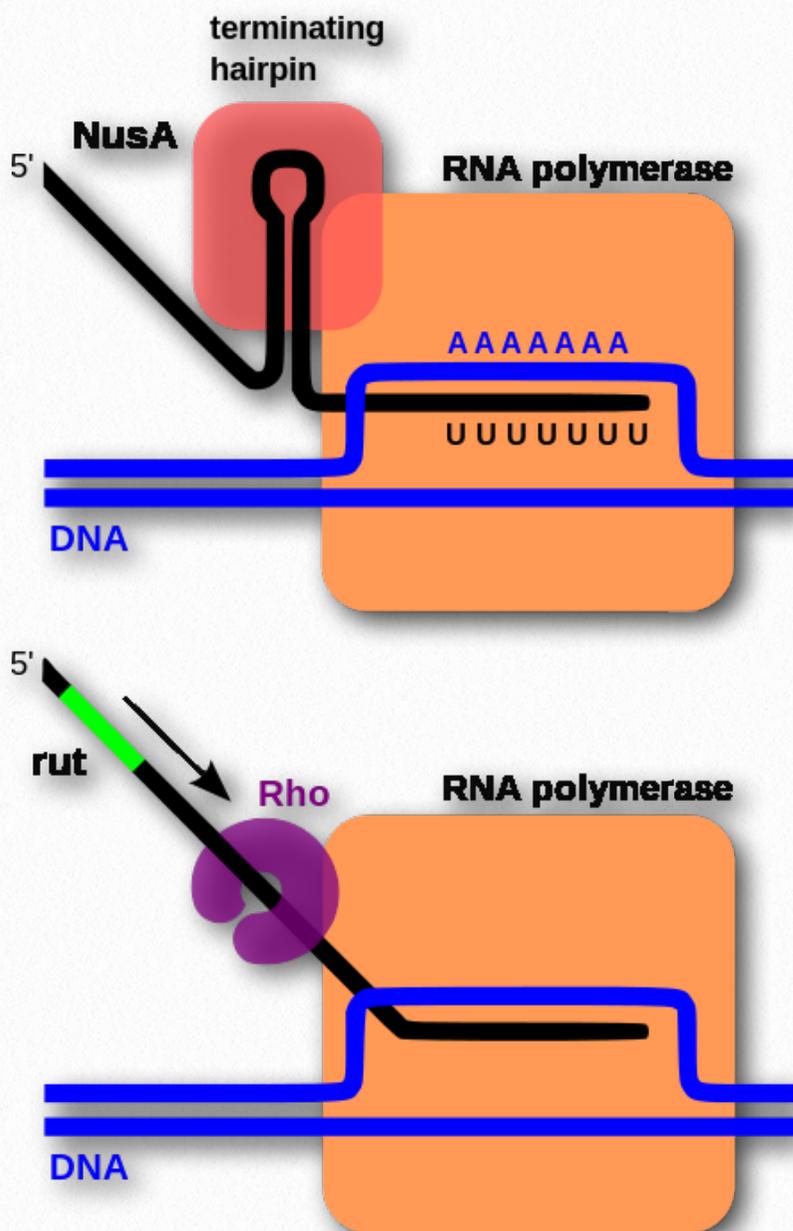
## Intrinsic terminators

In intrinsic terminators, this sequence in the RNA has self-complementary regions that can base-pair with each other to form a hairpin structure that contains a GC-rich run in the "stem" of the hairpin. This hairpin is fol-



**Figure 7.61 - Promoter and Terminator sequences determine where transcription starts and ends.**

lowed by a single-stranded region that is rich in U's (Figure 7.62). The secondary structure formed by the folding of the end of the RNA into the hairpin causes the RNA polymerase to pause. Meanwhile, the run of U's at the end of the hairpin permits the RNA-DNA hybrid in this region to come apart, because the base-pairing between A's in the DNA template and the U's in the RNA is relatively weak. This allows the transcript to be



**Figure 7.62 Transcription termination by intrinsic (top) and rho-dependent (bottom) mechanisms**

Wikipedia

released from the DNA template and from the RNA polymerase.

### Rho-dependent termination

In the case of rho-dependent termination, an additional protein factor, rho, is necessary. Rho is a helicase that can separate the transcript from the template it is paired with. As in intrinsic termination, rho-dependent termination requires the formation of a hairpin structure in the RNA that causes pausing of the RNA polymerase. Meanwhile, rho binds to a region of the transcript called the rho utilization site (rut) and moves along the RNA till it reaches the paused RNA polymerase. It then acts on the RNA-DNA hybrid, releasing the transcript from the template.

### Coupled transcription and translation

In prokaryotes, which lack a nucleus, the DNA is not separated from the rest of the cell in a separate compartment, so the mRNA is immediately available to the translation machinery, as the transcript is coming off the template DNA. Indeed, in prokaryotic cells, translation of the mRNA can begin before the entire gene has been transcribed. Ribosomes can assemble at the 5' end of the transcript, as it is displaced from the template, while the 3' end of the gene is still being copied. The lag time between transcrip-

tion and translation is thus, very short in prokaryotes.

## Transcription in eukaryotes

Although the process of RNA synthesis is the same in eukaryotes as in prokaryotes, there are some additional considerations in eukaryotes. One is that in eukaryotes, the DNA template exists as chromatin, where the DNA is tightly associated with histones and other proteins. The "packaging" of the DNA must therefore be opened up to allow the RNA polymerase access to the template in the region to be transcribed (Figure 7.63). The restructuring of chromatin to allow access to regions of DNA is thus an important factor in determining which genes are expressed.

## Multiple RNA polymerases

A second difference is that eukaryotes have multiple RNA polymerases, not just one as in bacterial cells. The different eukaryotic polymerases transcribe different classes of genes. For example, RNA polymerase I transcribes the ribosomal RNA genes, while RNA polymerase III copies tRNA genes. The RNA polymerase we will focus on most is RNA polymerase II, which

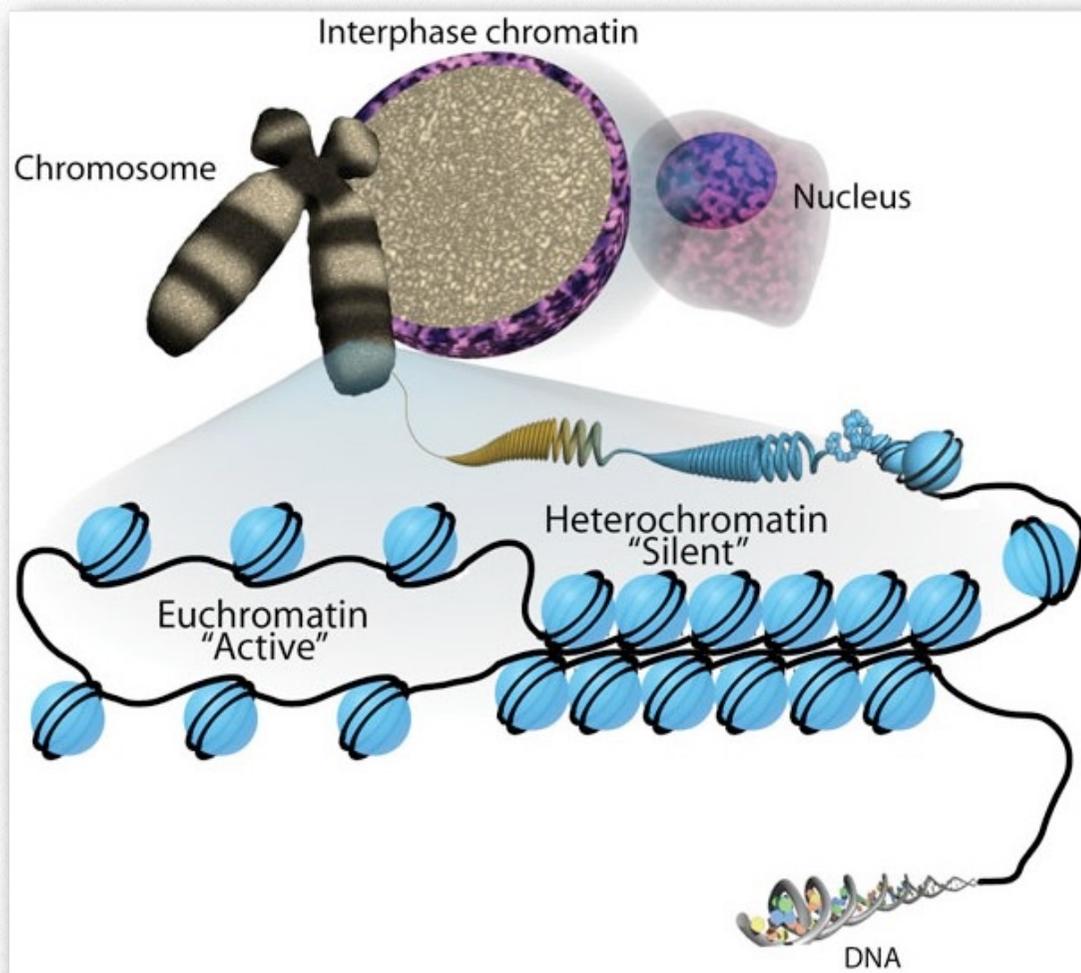


Figure 7.63 - Eukaryotic DNA is complexed with proteins in chromatin

Wikipedia

transcribes protein-coding genes to make mRNAs.

YouTube Lectures  
by Kevin  
[HERE & HERE](#)

All three eukaryotic RNA polymerases need additional proteins to help them get transcription started. In prokaryotes, RNA polymerase by itself can initiate transcription (the  $\sigma$  subunit is a subunit of the RNA polymerase, not an entirely separate protein). The additional proteins needed by eukaryotic RNA polymerases are referred to as transcription factors. We will see below that there are various categories of transcription factors.

## Transcription and translation are de-coupled

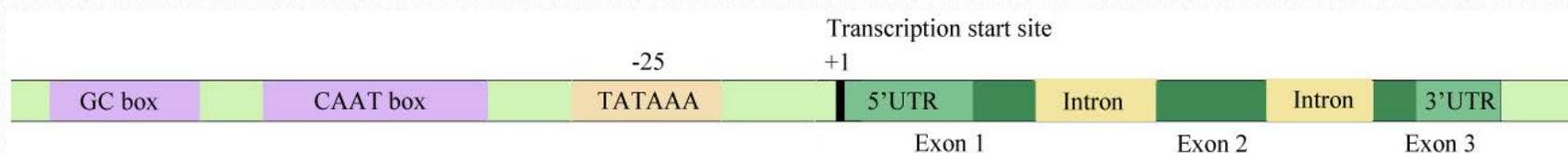
Finally, in eukaryotic cells, transcription is separated in space and time from translation. Transcription happens in the nucleus, and the RNAs produced are processed further before they are sent into the cytoplasm. Protein synthesis (translation) happens in the cytoplasm. As noted earlier, in prokaryotic cells, mRNAs can be translated as they are coming off the DNA template, and because there is no nuclear envelope, transcription and protein synthesis occur in a single cellular compartment. A representative eukaryotic gene, depicted in [Figure 7.64](#) shows that transcription starts some 25 bp downstream of the TATA box, and creates a transcript that begins with a 5' untranslated region (5'UTR) followed by the coding region which may include multiple introns and ending in a 3' untranslated region or 3'UTR ([Figure 7.64](#)). As detailed below, the initial transcript is further processed before it is used.

## Eukaryotic promoters

Like genes in prokaryotes, eukaryotic genes also have promoters that determine

where transcription will begin. As with prokaryotes, there are specific sequences in the promoter regions that are recognized and bound by proteins involved in the initiation of transcription. We will focus primarily on the genes encoding proteins that are transcribed by RNA polymerase II. Such promoters commonly have a TATA box, a sequence similar to the -10 sequence in prokaryotic promoters. The TATA box is a sequence about 25-35 basepairs upstream of the start of transcription (+1). (Some eukaryotic promoters lack TATA boxes, and have, instead, other recognition sequences, known as DPE, or downstream promoter elements.) Interestingly, the TATA box is not directly recognized and bound by RNA polymerase II. Instead, this sequence is bound by other proteins that, together with the RNA polymerase, form the transcription initiation complex.

Eukaryotic promoters also have, in addition, several other short stretches of sequences, that affect transcription, within about 100 to 200 base-pairs upstream of the transcription start site. These sequences, which are sometimes called upstream elements or



**Figure 7.64 - Region surrounding the transcriptional start site in eukaryotic DNA**

promoter-proximal upstream elements, are bound by activator proteins that interact with the transcription complex that forms at the TATA box. Examples of such upstream elements are the CAAT box and the GC box (Figure 7.64).

## Making transcripts in eukaryotes

We noted earlier that all eukaryotic RNA polymerases need additional proteins to bind promoters and start transcription. The proteins that help eukaryotic RNA polymerases find promoter sites and initiate RNA synthesis are termed general transcription factors. We will focus on the transcription factors that assist RNA polymerase II, the enzyme that transcribes protein-coding genes. These transcription factors are named TFIIA, TFIIB and so on (TF=transcription factor, II=RNA polymerase II, and the letters distinguish individual transcription factors).

Transcription by RNA polymerase II requires the general transcription factors and the RNA polymerase to form a complex, at the TATA box, called the basal transcription complex or transcription initiation complex (Figure 7.65).

This is the minimum requirement for any gene to be transcribed. The first step in the formation of this complex is the binding of the TATA box by a transcription factor, TFIID. TFIID is made up of several proteins, one of which is called the TATA Binding Protein or TBP. Binding of the TBP causes the DNA to bend at this spot and take on a structure that

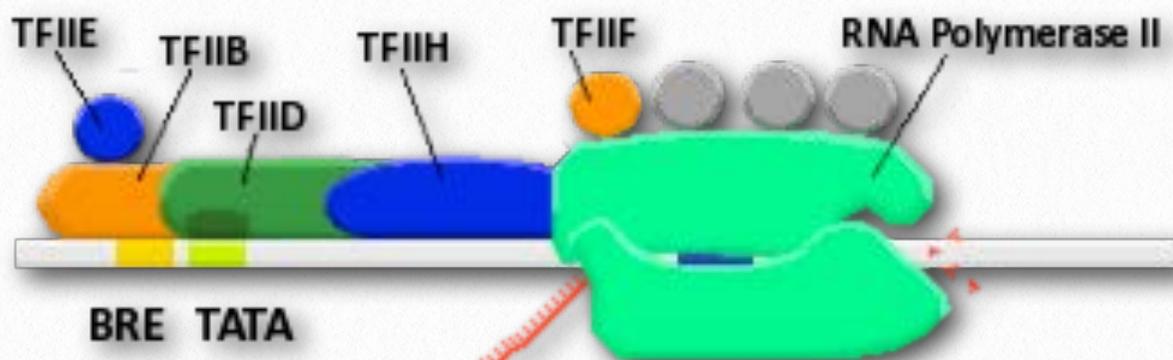


Figure 7.65 - Transcription pre-initiation complex in eukaryotes

Wikipedia

is suitable for the binding of additional transcription factors and RNA polymerase.

Interestingly, the binding of the TBP is a necessary step in forming a transcription initiation complex even when the promoter lacks a TATA box. The order of binding of additional proteins after binding of the TBP, as determined through *in vitro* experiments, appears to be TFIIB, followed by TFIIF and RNA polymerase II, then TFIIE. The final step in the assembly of the basal transcription complex is the binding of a general transcription factor called TFIIH. Some evidence suggests that following the binding of the TBP to DNA, the rest of the proteins in the initiation complex may assemble as a very large

# The Student's Guide to a Perfect Transcript

Pol II's so smart, now let me see  
It makes a transcript, one, two, three,  
From a billion base pairs can it find  
The one promoter it must bind?

I need some help, I hear it plead,  
This DNA I cannot read  
I cannot see where I must start  
'Til TFIIs have done their part.

When TATA's bound by TBP  
That sends a signal out, you see  
Once DNA has made a bend  
More TFIIs will soon attend.

When B arrives, it clears a place  
For RNA polymerase  
And F and E soon join the fun  
TFIIH, the final one.

The last one is a special case  
It moonlights as a helicase  
And sends Pol II upon its way  
To make a brand new RNA.

How does it do that, you may ask  
A phosphate does the crucial task  
Added on to CTD,  
That's how the H sets Pol II free.

And Pol II goes its merry way  
There soon will be some RNA  
Then introns must be all excised  
As RNA is capped and spliced.

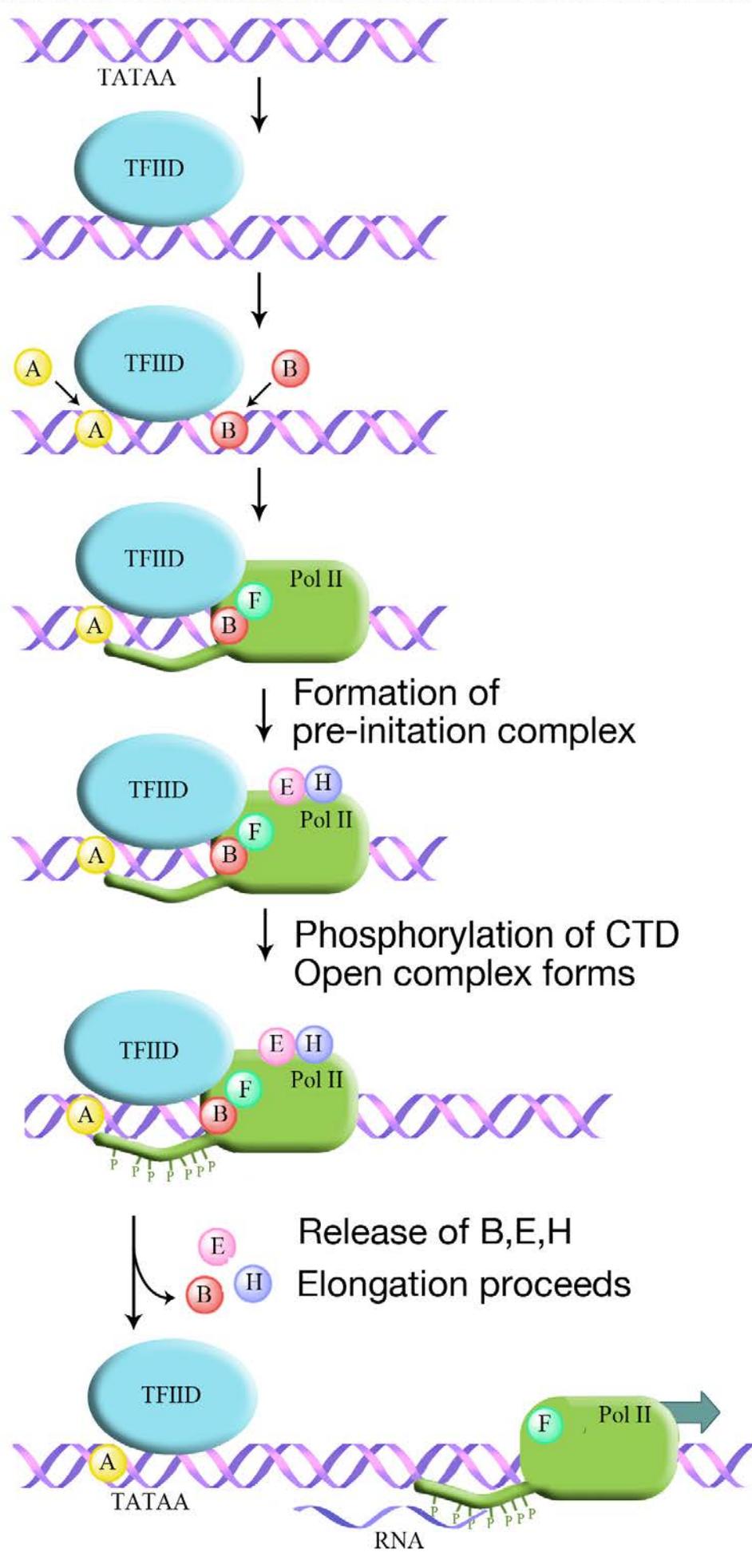
Then PolyA polymerase  
Will add a couple hundred As  
For my own transcript, I can say  
That all I want is one more A

*Verse by Indira Rajagopal*

complex that then binds directly to the DNA. In any case, the presence of all of these general transcription factors and RNA Polymerase II bound at the promoter is necessary for the initiation of transcription.

As in prokaryotic transcription, once the RNA polymerase binds, it can begin to assemble a short stretch of RNA. This must be fol-

lowed by promoter clearance, in order to move down the template and elongate the transcript. This requires the action of TFIIH. TFIIH is a multifunctional protein that has helicase activity (i.e., it is capable of opening up a DNA double helix) as well as kinase activity. The kinase activity of TFIIH adds phosphates onto the C-terminal domain (CTD) of the RNA polymerase II. This



**Figure 7.66 - Assembly of the basal transcription complex and transcription initiation**

phosphorylation appears to be the signal that releases the RNA polymerase from the basal transcription complex and allows it to move forward on the template, building the new RNA as it goes (Figure 7.66).

Termination of transcription is not as well understood as it is in prokaryotes. Termination does not occur at a fixed distance from the 3' end of mature RNAs. Rather, it seems to occur hand in hand with the processing of the 3' end of the primary transcript. The polyadenylation signal in the 3' untranslated region of the transcript appears to play a role in RNA polymerase pausing, and subsequent release of the completed primary transcript. Recognition of the polyadenylation signal triggers the binding of proteins involved in 3' end processing and termination.

**YouTube Lectures  
by Kevin  
[HERE](#) & [HERE](#)**

Graphic images in this book were products of the work of several talented students.  
Links to their Web pages are below

Click [HERE](#) for  
Aleia Kim's  
Web Page

Click [HERE](#) for  
Pehr Jacobson's  
Web Page

Click [HERE](#) for  
Penelope Irving's  
Web Page

Click [HERE](#) for  
Martha Baker's  
Web Page

Problem set related to this section [HERE](#)

Point by Point summary of this section [HERE](#)

To get a certificate for mastering this section of the book, click [HERE](#)

Kevin Ahern's free iTunes U Courses - [Basic](#) / [Med School](#) / [Advanced](#)

Biochemistry Free & Easy (our other book) [HERE](#) / [Facebook Page](#)

Kevin and Indira's Guide to Getting into Medical School - [iTunes U Course](#) / [Book](#)

To see Kevin Ahern's OSU ecampus courses - [BB 350](#) / [BB 450](#) / [BB 451](#)

To register for Kevin Ahern's OSU ecampus courses - [BB 350](#) / [BB 450](#) / [BB 451](#)

Biochemistry Free For All [Facebook Page](#) (please like us)

Kevin Ahern's [Web Page](#) / [Facebook Page](#) / Taralyn Tan's [Web Page](#)

Kevin Ahern's free downloads [HERE](#)

OSU's Biochemistry/Biophysics program [HERE](#)

OSU's College of Science [HERE](#)

Oregon State University [HERE](#)

Email [Kevin Ahern](#) / [Indira Rajagopal](#) / [Taralyn Tan](#)

# Central Dogma Zen

To the tune of "*Those Were the Days, My Friend*"

**Metabolic Melodies** Website [HERE](#)

Once upon a time a cell decided  
The time was ripe for it to split in two  
Had to copy cellular instructions  
For the daughter cell would need them too.

Bring in a helicase  
Unzip the DNAs  
To ease the stress a gyrase joins the fray  
Strands must be held apart  
SSBs do their part  
And primase builds a primer RNA.

Sliding clamp comes in behind clamp loader  
dNTPs floating all around  
In the wings a replicase is waiting  
For the chance to start another round.

Polymerase, my friend  
Starts at the 3' end  
It puts a T across from every A  
A G across from C  
Perfect simplicity  
The leading strand is made in just this way.

The lagging strand is made in little pieces  
Okazaki fragments, you recall  
Pol I fills the gaps that lie between them  
Ligase comes in next and joins them all.

Blueprints can't have mistakes  
That's why polymerase  
Corrects its work  
With exonuclease  
Proofreading one by one  
Till all its work is done  
Hurray for D-N-A polymerase!

An organism's cellular construction  
With blueprints for the things they have to do  
Requires converting DNA instructions  
To ribopolymers, oh yes it's true

Because they've been bestowed  
With a genetic code  
The RNAs provide the cell with means  
To link amino A's  
In most directed ways  
Inside the protein-making cell machines

If "*coli*" cells don't have galactosidase  
And lactose should appear inside its food  
The lac repressor leaves the operator  
'Cause otherwise metabolism's screwed

Polymerase unwinds  
The DNAs it binds  
Adjacent to the start site where it docks  
Unravels A's and T's  
With such amazing ease  
At the promoter's little TATA box

The process moves along without much trouble  
While making RNA inside the cell  
It all occurs inside transcription bubbles  
Where bases get linked anti-parallel

mRNA then roams  
To find some ribosomes  
Subunits large and small bind near the end  
The A-U-G's in place  
Inside the P site space  
Initiation you can comprehend

The mechanism shifts to elongation  
Proceeding by three bases at a stretch  
A GTP's required for translocation  
Advancing 5 to 3 the whole complex

The process moves anon  
Until a stop codon  
Arrives and causes movement to suspend  
Translation has to cease  
A peptide gets released  
And we have reached the central dogma's end

Recording by Tim Karplus  
Lyrics by Indira Rajagopal and Kevin Ahern

# Transcription

To the tune of "Frosty the Snowman"

**Metabolic Melodies** Website [HERE](#)

Phos-pho-di-esters  
Are the bonds of RNA  
That support a ribopolymer  
Made of G,C,U and A

The RNA polymerase  
Binds to a TATA box  
And copies from the template strand  
All the along the way it walks

IN-i-ti-a-tion  
Of transcription thus proceeds  
From the closed to open complexing  
In the DNA it reads

The sigma factor gets released  
Its work is over fast  
Polymerase can then advance  
After this step has been passed

In elongation  
The polymerizing spree  
Moves along the way in fits and starts  
Synthesizing five to three

The RNA is floppy and  
It dangles from one end  
Oh that's a most important thing  
For you to comprehend

Then termination  
Fin-ish-ES the RNAs  
Thanks to protein rho or hairpin forms  
That release polymerase

So this describes transcription's steps  
In three part harmonies  
Here's hoping with this melody  
You can learn it all with ease  
*fade*

Recorded by David Simmons  
Lyrics by Kevin Ahern

# Chromatin

To the tune of "Sunshine on My Shoulders"

**Metabolic Melodies** Website [HERE](#)

Charges of the lysines bind to phosphates  
Minus charges cling to plusses tight  
Chromatin assembly is essential  
Eukaryotic cells must get it right

Cells are tiny micro-scaled enclosures  
With nuclei tucked deep in their insides  
That's the place amazingly enough that  
Seven feet of DNA resides

H2a and H2b have lysines  
That's how they get charges don't forget  
Paired with H's three and four they make up  
A chromatin core's octameric set

These get organized in higher orders  
Changing with the cycles of the cell  
Denser packing going through mitosis  
At other times the structures simply swell

So because of all the hyperpacking  
Nuclei can hold entire genomes  
Thank the histones spooling DNA for  
Physically downsizing chromosomes

*Recording by David Simmons  
Lyrics by Kevin Ahern*