

MBB 206

TOPIC#1: Molecular Aspects of Transcription

Gene expression is a topic that includes: transcription, post-transcriptional processing and translation. Many points of regulation occur at these 3 steps for various reasons. My lectures will cover gene transcription. The topics we will cover in this first section are:

- proks vs. euks
- polymerase subunit composition
- Pol II and the CTD
- kinetics of transcription initiation
- the process of transcription
- basic aspects of promoter structure

Compare Genome Size

The increase in size and complexity of eukaryotic genomes requires a larger, more complex set of RNA polymerases and subunits of polymerases or as the size of the organism and complexity of the genome increases - so does the number of factors used to carry out transcription but

****the elegance of these mechanisms are the same in prokaryotes vs. eukaryotes*

Comparison of Genomes in the three Animal Kingdoms

Proks

- no nucleus
- polycistronic
- no introns
- translation begins before transcription ends
- **1 polymerase**
- **~5,000 genes**

Archaea

- no nucleus
- polycistronic
- no introns
- translation begins before transcription ends
- **1 polymerase**, but many subunits like euk pol.s
- **~ 2,000 genes**

Euks

- have a nucleus & nucleolus
- monocistronic
- genes have introns, **splicing**
- mRNA is translated in cytoplasm, mRNA must be exported
- **4 (5?) RNA polymerases**
4 classes of genes
- **~30,000 genes**

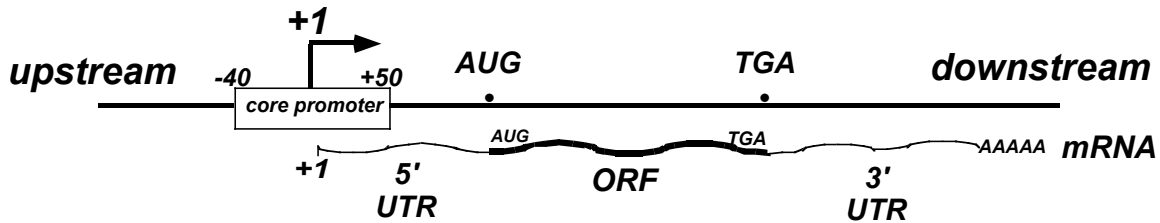
**RECENT HISTORY OF MOLECULAR BIOLOGY AND EVENTS IN UNDERSTANDING
EUKARYOTIC GENE EXPRESSION**

- 1972** first eukaryotic gene cloned in *E. coli* (from mammalian SV40 virus)
- 1975** Dr. Southern performed the first DNA-DNA hybridization on a membrane (referred to as a Southern blot)
- 1976** first cDNA made, from rabbit β -globin gene
- 1977** exon/intron gene structure first discovered
- 1977** DNA sequencing technology (called Maxam Gilbert, chain termination sequencing)
- late 1970's** Dr. Michael Smith (UBC, Vancouver BC) – invented useful *in vitro* mutagenesis techniques

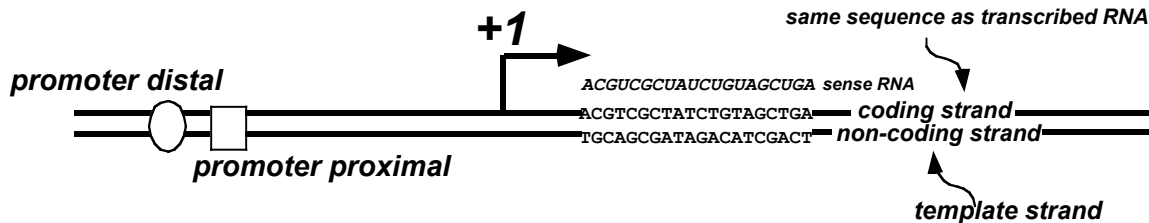
BASIC FEATURES OF A GENE AND ITS PROMOTER

Dictionary of Cell and Molecular Biology, 3rd Edition: Originally defined as the physical unit of heredity but the meaning has changed with increasing knowledge. It is probably best defined as the unit of inheritance that occupies a specific locus on a chromosome.

Molecular Cell Biology (Lodish et al): the entire nucleic sequence that is necessary for the synthesis of a functional polypeptide or RNA....includes all DNA sequences required for synthesis of a particular transcript (e.g., control regions as well as coding regions).



AUG: translation start codon
 TGA: translation stop codon
 UTR: untranslated region



Core promoter elements:

- bind to and control assembly of the RNA polymerase preinitiation complex
- position the transcription start site and control the directionality of transcription
- respond to nearby (proximal) or distal activators and repressors
- in most cases, core promoters do not play a direct role in regulated transcription
- support low or "basal" levels of transcription, the level being dictated by DNA sequences in the core promoter elements. In other words, core promoters determine frequency of transcription initiation.

RNA Polymerases

To start a discussion about transcription, we start with the business end of the whole process - the RNA polymerase - the protein complex that "reads" DNA sequence and synthesizes an RNA strand.

- RNA polymerases are very large complexes of proteins
- the best studied RNA polymerase is the *E. coli* RNAP
- many features of RNA polymerization are conserved
 - a. synthesize RNA in the 5' to 3' direction
 - b. require a DNA template
 - c. can synthesize strands *de novo* (no primer required)
 - d. they are processive

It is most instructive to start with a description of *E. coli* RNAP and then work into the eukaryotic polymerases, why?

- RNAP has a limited number of subunits
- large amounts of the enzyme can be purified, or you can buy pure RNAP

***E. Coli* RNA polymerase (RNAP) Composition:**

<u>Name</u>	<u>Number</u>	<u>Mass</u>	<u>Function</u>
α	2	37 kD	activation, nucleates assembly of RNAP
β	1	150 kD	Catalytic, binds NTPs, forms phosphodiester bonds. Mg^{+2} binding
β'	1	155 kD	Binds non-transcribed strand, termination. Mg^{+2} binding
σ	1	varied*	Recognizes specific promoter sequences, also participates in transcription activation

total mass ~ 420 kd

- there are known structures for individual polymerase subunits of eukaryotic RNAP, but recently Roger Kornberg and colleagues solved the crystal structure for a "halted" RNA PolII (*S. cerevisiae*; Cramer et al. 2000. *Science* 288:640-649). Seth Darst recently solved the crystal structure

for purified *E. coli* RNAP For very nice pictures of the EM structure of RNAP see: (Opalka et al., Proc Natl Acad Sci U S A. 2000 Jan 18;97(2):617-22) Also, there are EM high resolution pictures that you can view at the following site.

<http://www.rockefeller.edu/labheads/darst/structures.htm>

$\alpha_2\beta\beta'$ = **CORE polymerase**: can bind to DNA, but OFF rate is $\sim 10^4$ > than **Holo** polymerase (core + σ). Only capable of initiating transcription randomly (e.g. from nicks and gaps)

Looks like a bean with a hole in it, or a three-fingered hand. Due to my limited artistic ability, this is how the RNAP will be drawn in class:



* σ : there are multiple σ 's in the cell

1. each are sequence-specific DNA binding proteins
2. reduce non-specific DNA binding of the RNAP
3. increase specific binding to promoters

	-35	spacer	-10	Class of genes
σ^{70}	TTGACA	16-18	TATAAT	"housekeeping" (95% of genes)
σ^S			CTATACT	Stationary phase
σ^{32}	CTTGAAA	11-16	CCCATnT	Heatshock/stress
σ^F	WGCATA	14-15	GGnRAYAMTW	Flagella/chemotaxis
σ^E	GACTT	16-17	TCTRA	Extracellular functions
σ^{fecl}				Extracellular functions
σ^{54}	TGGCAC (-24)	5	TTGCW (-12)	Nitrogen metabolism

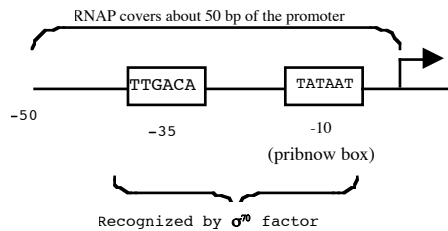
From: FEMS Microbiology Reviews 22:127-150 (1998)

σ -70 most common and best studied

σ -38 recognizes promoters important for maintenance during stationary phase or growth. At this phase the supercoil density of the genome is low. σ 38 has a high affinity for relaxed DNA, not highly supercoiled DNA

σ -54 needed for regulation of operons for nitrogen metabolism. Represents a whole different, second family of sigmas and widely distributed in prokaryotes. It is distinct structurally and functionally from the sigma70 family, and its mode of transcription looks more eukaryotic-like. σ -54 promoters can only be activated never repressed because there is no basal transcription in the absence of activators, only a closed complex.

**σ 's direct different gene activity at different promoters.
Think of them as promoter selectivity factors.**



how does RNAP recognize a promoter sequence? Atomic Force Microscopy confirms scanning. AFM showed scanning recognition mode in real time

Eukaryotic RNA polymerases

To isolate and identify all components necessary to initiate transcription, Bob Roeder and Bill Rutter discovered 3 RNA polymerase activities in eukaryotes (1969). Using DEAE sephadex chromatography, they designated different polymerase activities I, II, III by order of their elution with increasing NH_4SO_4 concentration.

1979, the first in vitro Pol II transcription in cell-free extracts (accurate transcription) was established.

1984, biochemists had purified 3 core eukaryotic RNA pol.s: Polymerase **I**, **II**, and **III**

NEW!

A new eukaryotic polymerase has been discovered: *Kravchenko et al. 2005. Nature 436:735-739*

This new polymerase is an alternatively spliced derivative of the mitochondrial RNA polymerase but doesn't have the mito-targeting signal and instead moves into the nucleus. It is resistant to alpha-amanitin and apparently transcribes mRNA for coding sequences. Nucleotide sequence alignment of the promoters of three gene targets highlights a loosely conserved sequence comprised of two unique elements. Too soon to know how essential and functional this polymerase is.

This brings the total number of known eukaryotic RNA polymerases to FIVE.

SIMILARITIES BETWEEN *E. coli* RNAP AND EUKARYOTIC POLYMERASES

- All Prok and Euk core polymerases can transcribe DNA if primed from a nicked or single-stranded template or from an annealed oligonucleotide. The polymerases bind to nicks and single-stranded regions in a non-specific manner.
- The 2 largest subunits of each polymerase are similar to the two largest subunits of prokaryotic subunits, β , β' . These contain portions of the catalytic site of RNA synthesis
- 2 subunits are similar to the prokaryotic α subunit
- there are many more subunits
- some subunits are shared by all 3 polymerases

Note: eukaryotic polymerases will be depicted like the bean-shaped coli RNAP

DIFFERENCES BETWEEN EUK AND PROK RNA POLYMERASES

➔ Unlike prokaryotic HOLO polymerases, eukaryotic "HOLO" polymerases are incapable of accurate, class specific promoter recognition. They cannot bind to the promoter.

- nuclear location (eg. RNA pol I in the nucleolus)
- chromatographic behavior
- salt requirements
- subunit composition
- * sensitivity to α -amanitin
- class of RNA they synthesize

Pol.	location	Type of RNA transcribed	α amanitin sens.*
I	nuc./nucleo.	rRNA (except for 5S RNA)	resistant
II	nucleus	hnRNA (i.e. pre-mRNA)	Very sensitive!
III	nucleus	Small RNAs such as tRNA, 5SRNA	Mod. sensitive
Mito	mitochondria	mitochondrial genome (diff. genetic code)	resistant
spRNAPIV	nucleus	Class? protein coding? promoters may have novel motifs	resistant

α amanitin: A toxin found in the poisonous deathcap mushroom, *Amanita phalloides*. This mushroom is considered to be the deadliest mushroom in the world. It likes to grow in California (Central/Northern region) under or around hardwoods, primarily oak trees. Its action is to potently inhibit RNA polymerase II transcription by binding to the β subunit and preventing the 3rd/4th nucleotide from being added to the RNA chain.

Recently Werner and Weinzierl succeeded in putting together an active Archae polymerase with purified recombinant subunits. 2002. *Molecular Cell* 10:635-646.

- how have subunits been defined?
 1. biochemical purification
 2. co-precipitation with monoclonal Antibodies against specific subunits or artificially engineered epitope tags (identify bands on a gel by molecular weight).

Structure of yeast RNA polymerase II at 5 Å resolution reference: Cell, Vol. 98, 799–810, September, 1999 "Yeast RNA Polymerase II at 5 Å Resolution" Jianhua Fu, Averell L. Gnat, David A. Bushnell, Grant J. Jensen, Nancy E. Thompson, Richard R. Burgess, Peter R. David, and Roger D. Kornberg. Now there is an even better structure of RNA polymerase (Westover et al. 2004. Cell 119:481-489).

Phosphorylation is important from initiation to elongation

why?: Ab specific for form A (Π_A) inhibits initiation, not elongation so elongating complexes have a phosphorylated CTD tail(Π_0).

Remember!

an important distinction between Prokaryotic and Eukaryotic (and Archaeal) Polymerases:

Unlike prok Holoenzyme, euk "HOLO" enzyme is incapable of accurate, class specific promoter recognition. It cannot bind to the promoter. What factors help eukaryotic polymerases find the right promoter?

Answer: GTF's

KINETIC STEPS OF TRANSCRIPTION INITIATION
--

The basic similarities between prok and euk systems, ie. basic kinetics properties are similar between all the polymerases. The following kinetic steps are true for prok and euk systems and have been confirmed/established experimentally.

STEPS:

1. Binding
2. Isomerization
3. Abortive initiation
4. Elongation
promoter clearance is a pre-requisite to elongation

PLEASE **KNOW** THESE!!!

Also, please **KNOW** these terms:

- A. closed complex
- B. open complex
- C. ternary complex
- D. elongation complex

Please **KNOW** which step(s) require ATP, and which are ATP-independent

We will revisit these steps in greater detail throughout my lectures as I describe different interesting examples of regulation of transcription

The steps for transcription are basically the same amongst proks and all three polymerases in eukaryotes. **What will differ is the mode of promoter recognition, and the methods of regulation.**

1. RNAP binding (holoenzyme) $K_B = 10^6 - 10^9 M^{-1}$

result: CLOSED COMPLEX

there is an intermediate step with a conformational change: $-\beta\beta'$ wrap around the DNA

initial binding is non-specific

this complex is VERY unstable ($t_{1/2} < 1$ second)

2. Isomerization: DNA template melts $K_2 = 10^{-3} - 10^{-1} \text{ sec}^{-1}$

VERY fast

result: OPEN COMPLEX, a single-stranded bubble in the duplex template forms

$\sim -10, +3$ around the start site of transcription

OPEN COMPLEX is *very* stable ($t_{1/2} = \text{hours}$), making this step essentially irreversible

Note that isomerization on eukaryotic promoters requires ATP whereas prokaryotic

isomerization does not (the exception is σ^{54} which also requires ATP).

3. Abortive initiation: first phosphodiester bond is formed \rightarrow dinucleotide = TERNARY COMPLEX. As drawn below:

after the dinucleotide is made, more rNTP's can be added up to 9

* at any time, the chain can be released and the polymerase remains bound at the promoter. It is estimated that 90% of RNAPs do not undergo escape commitment from strong viral promoters – so next to polymerase recruitment, promoter clearance may be the next significant rate-limiting step (!?; Kugel and Goodrich, 1998. PNAS 95:9232-9237)

Switching from Initiation to Elongation:

Recent results have highlighted that the actual switch (termed Promoter Clearance) is a highly regulated event reflected in the phosphorylation state of the CTD domain of RNA polymerase II. *Hypo*-phosphorylated RNAP is recruited to promoters, *hyper*-phosphorylated RNAP is found in elongation complexes. This is an extremely important difference that we will discuss several times in the first three topics – so it is important to remember.

4. Elongation (rate in vitro: 100-300 bases/min; rate in vivo: 1200-2000 bases/min.)

for RNAP, σ is released and the polymerase takes on the form of an ELONGATION COMPLEX. Here it is important to point out the differences between prokaryotes and eukaryotes. Elongating RNA polymerase II has a heavily phosphorylated CTD tail. Because bacterial RNAP does not have such a tail, phosphorylation modifications are not noted. We will

return to phosphorylation of CTD because it is so central to regulation of POLII gene expression and even post-transcription processes such as RNA capping and splicing! After a RNA chain of 9 nucleotides is created further polymerization requires PROMOTER CLEARANCE. Usual promoter clearance time is 1-2 secs, but this varies widely from promoter to promoter for reasons we don't understand very well. We will see later that promoter clearance can be a point at which regulation occurs.

?How do transcription factors regulate transcription initiation? By affecting any of these steps. Promoter binding (recruitment) is by far the most common mechanism known for Pol II, the polymerase that will receive the most attention in the following lectures.

How do polymerases get to their targets?

CORE polymerases can transcribe non-specifically. They **cannot**:

1. find specific promoters (bind to specific elements)
2. transcribe directionally
3. respond to regulators (ie. transcription activators)

HOLO prokaryotic polymerase (RNAP) can bind to a promoter, because sigma factor is present. What about eukaryotic polymerases?

Bob Roeder, Jim Manley and their colleagues used biochemical fractionation to identify protein complexes other than polymerase that were necessary to attain site- and direction-specific transcription initiation from a eukaryotic RNA polymerase II promoter (*Weil et al. 1979. Cell 18:469-484; Manley et al. 1980. PNAS 77:3855-3859*). Some of the most common promoters used were highly active viral promoters. For example the Adenovirus major late promoter with a strong TATA box consensus sequence was used to study RNA polymerase II transcription. Eight additional factors for RNA polymerase II were identified. More recently, others have used yeast genetics to identify factors necessary for correct RNA polymerase I transcription (eg. Nomura's group here at UC Irvine). Thus, protein biochemistry was the first tool used to analyze eukaryotic transcription *in vitro*, but genetics have proved an equally valuable and indispensable tool. Keep in mind the merits of combining BIOCHEMISTRY and GENETICS

BIOCHEMICAL FRACTIONATION – Roeder

Method: nuclear extracts over columns
Assay: **in vitro transcription**

In vitro transcription assays

RUN-OFF TRANSCRIPTION ASSAY

- DNA template (linearized)
 - Mg^{+2} , salt (usually KCl)
 - nuclear extract, or whole cell extract, or purified proteins
- preincubate template with buffers/salts/extract 10'-15'*

- NTP's (one of the four is α -³²P-radioactive and DO NOT USE dNTP's)
addition of nucleotides starts the transcription process
- SDS + EDTA + proteinase K, carrier RNA, sodium acetate
this buffer stops transcription immediately, destroys protein and prepares the sample for cleaning with phenol/chloroform and then for ethanol precipitation of DNA and RNA
- phenol/chloroform extraction
- ethanol »»» ppt nucleic acids
- analyze on denaturing polyacrylamide gel
- measure amount of transcript (specific size)


RUN-OFF METHOD USING G-LESS CASSETTES


- as above only the sequences downstream of the transcription start site **do not contain any Guanine** residues for a defined sequence length
 - at the end of this sequence is a G-rich sequence
 - NTP mix is missing GTP, therefore, transcription can proceed until the G-rich stretch of sequence is encountered
 - RNA Polymerase II stalls at the G-rich sequence due to the lack of substrate to include in the elongating RNA chain
 - reaction is terminated as above and analyzed on denaturing polyacrylamide gel

PRIMER EXTENSION TRANSCRIPTION ASSAY

- DNA template (usually circular or supercoiled)
- Mg⁺², salt (usually KCl)
- nuclear extract, or whole cell extract, or purified proteins
preincubate step
- NTP's (none are radioactive)
transcription proceeds
- Stop transcription and process RNA samples as for RUN-OFF procedure above
- hybridize radiolabeled specific **antisense** oligonucleotide to RNA samples in presence of high salt (~300mM KCl)
 - collect annealed complexes by ethanol precipitation
 - resuspend nucleic acid with buffers and **Reverse Transcriptase** and dNTP's
 - collect **cDNA** products by ethanol precipitation
 - analyze on denaturing polyacrylamide gel
 - measure amt of transcript (specific size)

 **primer extension can be used to map transcription start sites**

 Single round transcriptions are performed by adding the detergent Sarkosyl after preinitiation/open complex formation to 0.16% final concentration. This level of Sarkosyl inhibits pre-initiation formation but does not affect engaged and elongating polymerases. Thus transcription that has started when Sarkosyl was added can finish, but no new rounds of transcription can occur. In this way, we can examine the initial kinetic steps more cleanly.

 Run-OFF or Primer extension assays can be used to check fractions off a column to follow elution profiles of proteins that regulate transcription, or the presence of basal factors such as RNA polymerase. It was the Run-Off assay that Roeder used to isolate and characterize **General Transcription Factors (GTFs)** for RNA polymerase II transcription. He found that these factors were necessary to enable PolII to find and bind a promoter.

Proceed to the Willy/Kadonaga paper