Core Promoters

All promoters (review):

- provide nucleation site for polymerases
- establish directionality
- establish start site from a defined, conserved distance
- what is the easiest way to tell if a gene is transcribed by Pol I, II, III?

answer: α-amanitin

RNAP

RNA POL I

- rRNA
- genes in clusters, 5 clusters with ~40 genes per cluster, ~200 genes/haploid
- very little variation
- genes are highly transcribed

rRNA genes <1% genome
rRNA ~ 60% total cellular RNA

RNA POL II

- mRNA, snRNA
- most variable class of genes

RNA POL II (cont.)

RNA POL III

- 5s RNA, U6 RNA, ???

General Transcription Factors and Mediator

For an in-depth discussion of these complexes and their role in transcription, the review by Orphanides et al. is superb. *Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes Dev. 1996 Nov 1;10(21):2657-83.*

General Transcription Factors for RNA polymerase II

Purified RNA Pol I, II or III cannot transcribe properly
requires additional factors to:
- bind promoter
- transcribe in the right direction
- choose site accurately

Factors that recruit a polymerase, direct it and help it initiate are BASAL FACTORS or GENERAL TRANSCRIPTION FACTORS (GTF'S)

A note about basal transcription: under normal conditions, the minimal protein apparatus required for accurate transcription initiation consists of the GTF's and the subunits of pol II. This transcription is not dependent on the presence of regulatory molecules and is known as "basal" transcription. Transcription in a cell however, is a tightly regulated process. DNA packaged into chromatin results in general gene repression. Transcription must be triggered by activator proteins that bind to specific DNA sequences and induce the epxression of a gene or set of genes (a topic that is addressed later in the course). This is known as "activated" transcription. Thus, basal transcription does not occur in vivo, and ispurely an operational term defined by in vitro studies.

All GTF's have been identified and almost all polypeptides cloned:

GTF:	**TFIIA***	**TFIIB**	**TFIID**	**TFIIE**	**TFIIF**	**TFIIH**	**(TFIIK)**
# subunits	3	1	~13	1	2	~9	

* TFIIA not needed in highly purified systems with strong promoters

All GTF's are described below, only two in any detail (TFIID, TFIIH)

TFIID

The promoter selectivity factor

- only GTF capable of sequence-specific binding to DNA
- is the largest GTF

- TBP is most well characterized subunit of TFIID
- TFIID ~750 kD
- TBP ~28 kD
  - the DNA binding subunit binds to TATA
  - binds minor groove of DNA, bends ~90°-110°
  - has homology to σ
  - flat convex surface

TBP was very, very difficult to purify from mammalian tissues - instead it was first purified and identified/cloned from yeast. Why? yTBP exists primarily as a free, monomer form (no one really knows why there is this difference). The lesson here is that sometimes it's good to work on the same problem in several different model systems. TBP's from other organisms were cloned based on low stringency screening strategies using yTBP as a starting point.

Early studies showed that highly purified RNA Pol II is not capable of promoter recognition. The search for factors that could direct PolII to a promoter, could direct proper start site choice and direction of transcription and could establish a basal level of transcription lead to the identification of general transcription factors (GTFs). But what about activated transcription? TFIID stood out as the most likely target of activators: it was the 1st GTF to bind to the promoter and it could specifically recognize the TATA element. However, for some time, this model remained untested because TFIID was difficult to purify. It wasn't until the cloning of yeast TBP in 1989 that this model could be tested in vitro. As I have mentioned before, for some reason, most of the TBP in yeast is found as an isolated protein and not bound up in myriad complexes as you might find in mammalian cells. The cloning of TBP allowed the isolation of TFIID, because antibodies, and epitope tagged versions of TBP allowed co-immunoprecipitation of the much larger TFIID complex. Also, TBP antibody could be covalently attached to columns and nuclear extract applied, any specific TBP containing complex could be eluted off, and specific binding to TATA sequences used to assay the fractions. These technologies, alongside an in vitro transcription assay with a TATA-dependent promoter led to the purification of TFIID.

Reconstituted transcription with purified RNA polymerase and GTFs missing the TFIID fraction were then used to test the model that TFIID is the target of transcription activators. In some of the earliest assays, the potent transcription factor Sp1 was used – it was one of the first mammalian transcription factors to be cloned, and its two transcription activation domains mapped to two Q-rich stretches. Experiments similar to the following were carried out.

--> The **big** surprise is that TBP is necessary to recruit all polymerase complexes: I, II, and III, it is NOT specific to RNA polymerase II transcription. Mutation or deletion of TBP in yeast destroys all forms of transcription in the nucleus (as opposed to transcription in mitochondria).

--> the DNA binding domain isn't always used to bind TATA. It's the large, interacting surface that is most necessary.

in mammals: TBP rarely exists free in the cell:

SL1 - RNA Pol I
TFIID RNA Pol II
$SNAP_c$ RNA Pol II, and III
TFIIIB RNA Pol III (TBP + Brf + 90K protein)

In the case of RNA polymerase III, two other GTF's, TFIIIA and TFIIIC do the job of recognizing POLIII core promoter elements. After they bind to the promoter, TFIIIB is recruited, followed by RNA POLIII recruitment. In this case, TBP (which is part of the TFIIIB complex) does

not function to recognize a DNA sequence, nor does it work within a complex that binds first to the promoter. This is very different than how it works on RNA POL II promoters. However, remember that I showed a class of POLIII promoter that contains a TATA element? Interestingly, TFIIIB binds to this TATA motif and bends the promoter. So even within a single class of RNA polymerase, TBP performs different functions depending on the core promoter sequence.

TAF's

All the other proteins in TFIID other than TBP

Green. 2000. Trends Biochem. Sci. 25:59-63.

--> what is TFIID? 750 kD = TBP + ~9 proteins called **TAF**'s for (**T**BP **A**ssociated **F**actors).

TFIID-associated TAF's

250 kD	HMG box, bromodomains, serine kinase, Histone acetyl transferase
150 kD	binds downstream promoter regions
135 kD	
95 kD	WD-40 repeats
80 kD	Histone H4 similarity
55 kD	
31 kD	Histone H3 similarity
28 kD	
20 kD	Histone H2B similarity

--> in human or drosophila extracts, TAF's bind so tightly they must be stripped off TBP with urea
--> how were they identified?

1. purification of TFIID to "homogeneity" = **holo-TFIID**. Protein gel shows many bands. To assay for TFIID during purification, used a combination of TATA binding and in vitro transcription.

2. TAF's can mediate activation of some transcription factors in vitro.

3. TAF's mostly not necessary for basal transcription with a TATA box containing promoter. However, TAF's can promote (may even be necessary) for basal transcription of TATA-less promoters. Certain TAF's can contact basal promoter elements other than TATA box such as the Initiator element.

4. TAF's can also be called selectivity factors - a feature of prokaryotic **σ** subunits. * since TBP is shared between all 3 polymerases, distinct sets of TAF's must dictate the type of promoter at which a given TBP•TAF will function

5. One of the most notable features is that several TAF's look like, and fold like histones. Current models suggest that TAF's form a nucleosome-like structure that helps wrap the bent DNA and prevent bona fide nucleosomes from clamping down and repressing transcription. This remains to be tested.

--> we now know that specific transcription factor activators contact specific TAF's. These TAF's are specifically required to reconstitute activated transcription by these special factor in vitro.

Early on a model for the function of TAFs was proposed: TAFs are the major, if not exclusive targets of transcription activators. In other words, activation of transcription involves recruitment of the TFIID complex to DNA via activator-TAF interactions. This model implies that TAF's are essential for most/all transcription activation events.

Question: Which kinetic constant describes RNA polymerase recruitment?

If this model is true, then knock-out of various TAF genes should cause widespread and catastrophic effects on RNA polymerase II transcription.

➔ This model was derived from in vitro transcription with partially purified protein preparations. What would be a good way to test this model?
   ➔ **Answer:** test the model *in vivo*. Yeast is a good organism to use for this test.

K. Struhl — null TAF mutations in yeast --> lethal
Moqtaderi et al. 1996. Nature 383:188-191.
M. Green — conditional mutations --> not a rapid effect on gene expression, not many promoters affected
Walker et al. 1996. Nature 383:185-188

At the minority of promoters whose transcription was impaired, TAF dependence was shown to reside in sequences in the vicinity of the TATA box. More recent studies show that TAFs clearly play an important role in promoter recognition, augmenting the minimal sequence specificity of TBP. In other words, TAF's round out the **σ** function of TFIID
****** In summary, these experiments ended up disproving the model as a universally true model in vivo. Instead, another large multi-subunit complex called **mediator (see later in notes)**. Appears to play the communicator role between transcription regulatory proteins and RNA Polymerase II.

Another recent surprise about TBP and TFIID is that there are at least two alternative TBP's identified. These proteins do not bind to TATA boxes. One of these alternative TBPs, TRF (for TBP related factor) participates primarily in RNA polymerase III transcription in *Drosophila* and a small amount of RNA polymerase II transcription. Another TBP-like protein (TLP) has been identified. Knock-out of TLP in early frog embryogenesis wipes out transcription of only specific PolII genes. Thus, we are now faced with the exciting prospect that alternative TBP proteins may function as important selectivity factors.

The other GTF's

How to recruit a polymerase and promote initiation

TFIIA

37 kD
19 kD — antirepression — required for activation
13 kD

TFIIB

35 kD recruits pol II/TFIIF, start site selection

TFIIF

58 kD (RAP74) — stimulates elongation
26 kD (RAP30) — s homology, cryptic DNA-binding, prevents spurious initiation

---

TFIIE:

56 kD — promoter melting, recruits and modulates activity of TFIIH
34 kD

TFIIH

• The subunits can be separated into 2 groups: **core + kinase**

89 kD (ERCC3)	3'->5' helicase (essential for transcription), excision repair
80 kD (ERCC2)	5'->3' helicase (NOT essential for transcripton), excision repair
62 kD	excision repair
44 kD (SSL1)	excision repair
40 kD (cdk7)	**CTD kinase, part of the cdk7-activating kinase (CAK)**
37 kD (cyclin H)	**cdk7 partner**
34 kD	
32 kD (MAT-1)	CAK assembly factor

core:

XPB/ERCC3/Rad25, 89 kD, 3'-->5' helicase
XPD/ERCC2/Rad3, 80 kD, 5'-->3' helicase
SSL1, 44 kD, DNA binding, Zn finger

XPB and XPD are previously known proteins involved in disease

<u>Xeroderma Pigmentosa:</u>

- patients have extreme sensitivity to sunlight --> hi chance of skin cancer
- profound growth retardation
- brittle hair
- delayed psycho motor development

XPB and XPD known to be involved in NER (**n**ucleotide **e**xcision **r**epair)
Thus, TFIIH is involved in both transcription and NER.

--> purified TFIIH can rescue the repair deficiency of mammalian cell lines or yeast strains mutated in repair genes.

<u>Transcription</u>
need only RAD25 3'-->5' for open complex formation and promoter clearance.

<u>NER</u>
need RAD3 <u>and</u> RAD25 together this unwinds DNA in both directions to start the repair process

• you can get rid of the IIH requirement by making <u>heteroduplex</u> DNA around the start site.
Otherwise you need: ATP, IIH, IIE to get the 1st phosphodiester bond.

(IIE affects the helicase activities)

---

kinase: aka TFIIK

- CDK7 is an integral subunit

CDK7 is the catalytic subunit of CAK (CDK activating kinase) which activates CDK2, CDK4, cdc2 - all of these are needed for cell cycle control

<u>How does CAK/TFIIK work in transcription?</u>

2 hypotheses: **1.** TFIIK phosphorylates the CTD of pol II to facilitate promoter clearance
2. TFIIK phosphorylates cdc2, CDK4 and/or CDK2 which in turn phosphorylate the CTD or other transcription factors

***** phosphorylation is not needed for 1st phosphodiester bond

So what is TFIIH doing in NER?

• A process called "transcription-coupled repair": DNA damage in the transcribed strand of actively transcribed genes is removed at a faster rate than in the genome overall.

• yeast experiments suggest:

2 different forms of TFIIH are involved in the 2 different processes, but it is TFIIH as a whole that participates in repair.

• yeast NER repairosome has no CTD-kinase activity (lacks TFIIK), so:

extensive DNA damage might lead to a down-regulation of Pol II transcription by the inactivation or release of TFIIK. release of TFIIK might also serve as a signal for check point control --> apoptosis

The case for another GTF: P-TEFb

All the GTF's listed above are found to be necessary for transcription initiation and at least one of them for promoter clearance (TFIIH). But recently, a new factor, P-TEFb (for positive transcription elongation factor "b") has been identified and its role defined as highly important for elongation and perhaps even for promoter clearance. Like TFIIH, P-TEFb has a cyclin (any one of three "T" types of cyclin or cyclin K) and a cyclin dependent kinase subunit (CDK9). P-TEFb phosphorylates the CTD tail of RNAP (on serine2) and it can also phosphorylate and inactivate elongation inhibitors. Why is this important? We will see in Topic#2 that there is a very large multi-subunit complex bound tightly to the hypo-phosphorylated CTD tail. This complex is called Mediator and it functions as a go-between certain transcription regulator factors that bind to regulatory elements in enhancers and promoters and the basal machinery (RNAP and GTF's). In other words, some transcription factors bind to Mediator and in so doing they help recruit RNAP to promoters. Phosphorylation breaks the association between Mediator and the CTD and thus allows RNAP to escape the promoter and begin transcribing. Also, as you will learn in the future lectures, a phosphorylated CTD tail of RNAP is an attractive binding target for RNA processing machinery such as capping enymes, polyadenylation factors and even splicing machinery. Thus phosphorylation of the CTD is a hugely important event and marks the transition from recruitment and initiation at the promoter, to escape and transcription elongation and ultimately RNA processing.

Why isn't P-TEFb a formal GTF? First and formost, P-TEFb is not required for standard in vitro transcription initiation reactions. But perhaps it is also because there are differential requirements for this factor at different promoters *in vivo*. Some promoters need it, others don't. P-TEFb requirements will likely be specified by the dynamics of RNAP pausing at the early steps of promoter clearance.

If you would like to read more about P-TEFb, here is a recent review:

Garriga and Grana. 2004. Cellular control of gene expression by T-type cyclin/CDK9 complexes. Gene **337**:15-23.

How do pre-initiation complexes form?

1. Multi-step model

TFIID+ P --> [TFIID•P] + TFIIB --> [TFIID•P•TFIIB] + POLII•TFIIF -->

[TFIID•P•TFIIB•POLII•TFIIF] + TFIIE•TFIIH -->
[TFIID•P•TFIIB•POLII•TFIIF• TFIIE•TFIIH]

this model is derived using purified components (Tjian lab, Sharp lab) and looking at promoter binding in vitro and in gel shift assays. However additional interactions between PolII and IIB, IIE and IIH have been detected. Therefore, in vivo a pre-assembled complex containing all of these factors may be recruited to promoters. This leads to the second model:

2. Holoenzyme assembly model

asserts that PolII exists in solution complexed to many of the GTF's already and that transcription activators recruit the holo enzyme through prot•prot contacts

• so what is a holo-enzyme?
• why does phosphorylation of the CTD occur? why do you need phosphorylation for promoter clearance?
• in some highly purified systems, TAF's do not mediate the function of all the activators. For example the yeast transcription factor Gal4 cannot activate transcription in vitro with all the GTF's and a highly purified preparation of pol II

Mediator - the activator target within RNA polymerase II complexes

Roeder. 2000. Trends Biochem. Sci. 25:277-283.

Several complementary approaches led to a recent development in the composition of RNA pol II: **yeast genetics** and **biochemistry**

1. Genetics with yeast Gal4 activator: Gal4 is a transcription factor that activates genes involved in galactose metabolism. Plating yeast on galactose media induces the expression of genes that allow it to metabolize this sugar. Gal4 activates these genes and is essential for this process. Initially, researchers thought that Gal4 activates transcription by TFIID recruitment, specifically through several TAF interactions. However, Gal4 gene targets in yeast were not compromised when the TAFs were knocked out. So, how is the transcription factor Gal4 activating transcription if it's not working through GTF's or TAF's? Expression of Gal4 in these mutants is able to activate transcription just fine. What is the mechanism by which Gal4 is activating RNA polymerase II transcription? Use yeast genetics. Mutate the activation domain of Gal4 and look for suppressors of this mutation.

Himmelfarb et al. 1990. Gal11P: A yeast mutation that potentiates the effect of weak Gal4-derived activators. Cell ***63****, 1299-1309.*

Result:

when the Gal4 activation domain is gone, mutations in a protein named Gal11 allow Gal11 to bind directly to the DNA binding domain of Gal4.

Barberis et al. 1995. Cell ***81****, 359-368.*

What is Gal11, and why does direct binding to Gal4 restore activation? No one knew for a long time what Gal11 was. Only one thing was for sure, it wasn't a DNA binding protein. Much later, results from another laboratory (see **3** below) intersected nicely with this project and provided the answer.

2. In vitro transcription systems -

Flanagan PM, Kelleher RJ, Sayre MH, Tschochner H, Kornberg RD. A mediator required for activation of RNA polymerase II transcription in vitro. Nature. 1991 Apr 4;350(6317):436-8.

Several groups, most notably Roger Kornberg's group noticed that addition of a purified transcription activator to a crude transcription extract system would interfere with the ability of another added activator to stimulate RNA polymerase II transcription. This process (now termed "squelching") is predicated on the notion that both activators compete for the same target protein or complex to induce activation, and this target must be in limiting amounts. At first the common target was thought to be RNA polymerase II itself or one or more of the GTFs. Adding back purified, excess polymerase or GTFs, didn't relieve the squelching. Fractions from extracts that could restore transcription and relieve the squelching were isolated and a large, mega-complex was purified and Kornberg named it "Mediator". He hypothesized that this complex served as an adaptor between activator proteins and the basal transcription machinery.

3. Yeast genetic experiments with RNA polymerase II CTD domain.

Nonet and Young. 1989. Genetics 123: 715-724

The final complementary set of experiments, was to cut off half of the CTD domain of the RNA pol II's β' subunit. This severely compromises transcription and yeast growth especially at colder temperatures (cold-sensitive phenotypes), but suppressors of this drastic mutation were isolated. There were multiple suppressors isolated and one of them was the gene that codes for Gal11. Other mutations

were in novel genes and Young called SRB, for Suppressors of RNA Polymerase B. All of the SRBs and Gal11 were found to be subunits of a large complex. Further characterization found that Kornberg's purified Mediator complex and the complex of SRB/Gal11 (and other proteins not mentioned here) are one and the same complex. Suppressor mutations in SRB's caused the proteins to bind more tightly to each other, or to the CTD tail of the RNA PolII β'-like subunit. Obviously yeast were adapting to the truncation mutation by strengthening the complex.

So how to explain the Gal4 data? We now know that Gal4 activates transcription by binding to, and recruiting Mediator/RNA polymerase II to its target promoters. In normal yeast, the part of Gal4 that binds to Mediator is the transcription activation domain. So, in the original experiment when the activation domain of Gal4 was removed, most yeast could not survive plating on galactose media. Mutant strains that could survive and use galactose as a carbon source in spite of a severely crippled Gal4 protein, were those that were lucky to have generated a gain-of-function mutation in a component of Mediator. This gain-of-function mutation was in the gene for Gal11, and the mutation caused Gal11 to bind to the DNA binding domain of Gal4. Thus recruitment of Mediator was restored.

4. Mediator purification

Thompson, et al. 1993. Cell 73: 1361-1375.
Koleske and Young. 1994. Nature 368: 466-469.
Kim et al. 1994. Cell 77:4599-599.

Ultimately, the herculean task of purifying this complex was completed and 16 proteins present in equimolar amounts on a gel were identified. All 16 genes that code for these subunits have been identified and characterized, and crude EM-derived structures have been determined. Also, mediator-like complexes have been purified from other systems and cell types, and the subunit composition is similar but not always identical.

*******Consistent with the genetics, Mediator complex binds tightly to the CTD domain of RNA POLII β'. In fact, one can purify holo-RNA polymerase II and treat the purified protein with an antibody to the CTD domain. The antibody is able to disrupt Mediator interaction with the CTD, and all 16 polypeptides are eluted. Furthermore, Mediator does not bind to phosphorylated CTD.

Question: why is it important for Mediator to have a low affinity for phospho-CTD?

Here are the current controversies/unknowns/problems:

• how does mediator actually work?

Some think by recruitment, others have evidence that it affects steps of initiation beyond that (eg. Mediator activates TFIIH kinase activity, what kinetic step would be affected?). Clearly, much needs to be learned about the molecular mechanism. The difficulty of studying a large multi-subunit complex *in vitro* is made even more difficult by the certain need to perform experiments on chromatin-assembled templates.

• Is there one Mediator or several?

A wide variety of Mediator complexes have been identified and the different complexes have similarities but also a few differences. While most of the complexes are similar to one another, the differences tend to be that groups of subunits are missing and this group of proteins can be found in extracts as their own multi-subunit entity. This observation has led several in the field to propose that Mediator is modular. Subsets of complexes can come and go and change the specificity of the Mega-complex. To date, 7 different mediator complexes have been described. (For now, we will not focus on

any one subunit, or sub-group of Mediator. Later when we discuss chromatin and its modification, we will study one or two Mediator components.)

Malik and Roeder. 2000. Trend in Biochem. Sci. 25: 277-283