

MBB206

3rd Topic: Activators and *cis*-elements

Cis-acting Elements near promoter

- functional definition
- how to find them
- how to analyze them

Remember that CORE polymerases can only transcribe non-specifically. They cannot:

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

•Basic promoter elements close to the start site of transcription encode binding sites and recognition elements for GTF's and other basic factors. Together they recruit the specific polymerase, the polarity of their interaction with DNA defines the direction that transcription will proceed.

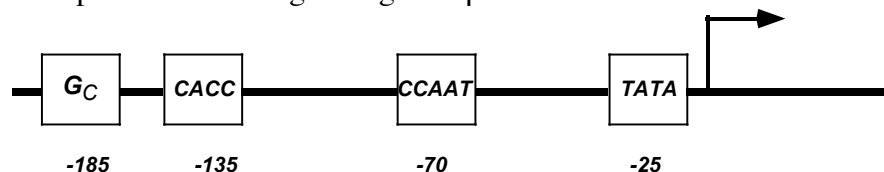
•Some elements are recognized by transcription activators that, if present and active, can elevate transcription rates above baseline.

Promoter: 2 Basic Regions

i. core element: start site control, direction control: recruits basal factors, RNA pol. and associated factors. These basic elements were presented before when we described the general promoter architectures for the three basic polymerases. Examples here are the TATA box, Initiator, DPE, and BRE (the NEW TFIIB element for RNA pol II). Remember the core elements for RNA pol I, and for RNA pol III.

ii. upstream elements within hundreds of base pairs (typically ~100-200 bp). Although upstream elements cannot be moved large distances, they show remarkable flexibility in that they can be moved around and sometimes flipped in orientation and they can still work. **There is flexibility.**

How are promoters analyzed for regulatory sites near the basal promoter? Look at a RNA polymerase II promoter: hemoglobin gene - β :



There are several common elements found in promoters. The elements are recognized by transcription factors. For example two of them in this promoter are:

- | | |
|--------|--|
| CCAAT | (C/EBP) |
| GC box | (Sp1) - Zn finger type DNA binding protein |

- these elements can sometimes work in both orientations (although the closer they are in to the promoter, the more constrained their position is)
- these elements are recognized by constitutively active proteins. Therefore these elements contribute to basal promoter activity when the gene is active (available for transcription).
- these elements work to increase the efficiency of the promoter - they mostly affect initiation (remember that this comprises MORE THAN ONE step).
- some of these elements are recognition sites for tissue specific transcription factors, thus upstream elements near the start site of transcription can establish **tissue specificity**. However, this is **often NOT** the case. Meaning, the basic promoter is promiscuous and able to work in many different cell types:

specificity:

some promoters are specific for:

- tissue type (OFF/ON)
- development stage
- temporal (eg. a certain stage in the cell cycle)
- environment (eg. signals, growth, DNA damage, nutrients, O₂)

Constitutive:

some genes need to be expressed in all cells at all times - these are "housekeeping genes", and their promoters are highly active and tend to be GC-rich. *In fact, ≥50% of the RNA polymerase II promoters in mammalian genomes are located in CpG islands - a fact exploited by the bioinformatic search algorithms designed to identify PolII promoters in genomic sequences.*

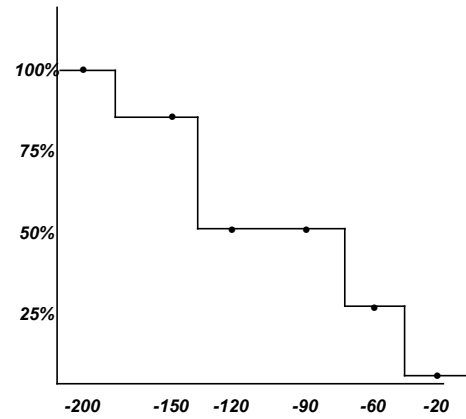
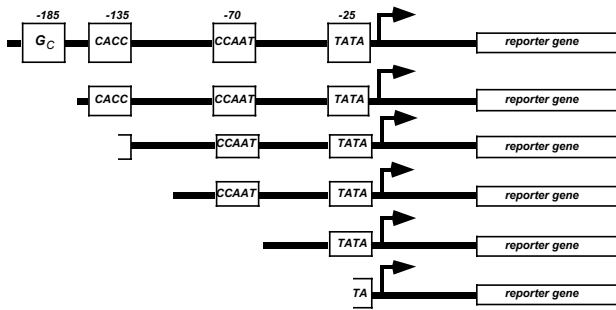
How do we define promoters experimentally?

1. Find the start site

There are many methods - but a widely used one is Primer Extension (see notes and figures for TOPIC #1). For this technique, you must have cloned genomic DNA (in/o/wo the promoter) and a cell source in which the gene is highly expressed.

2. Map boundaries

- a. need an assay for promoter activity
 - i. *in vitro* transcription with cell extracts
 - ii. *in vivo* transfections (transient)
- b. nested deletions



make 5' and 3' deletions. Either use convenient restriction enzyme sites or for a more systematic approach, use Bal 31 exonuclease digestion.

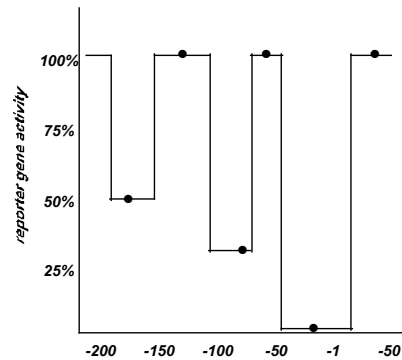
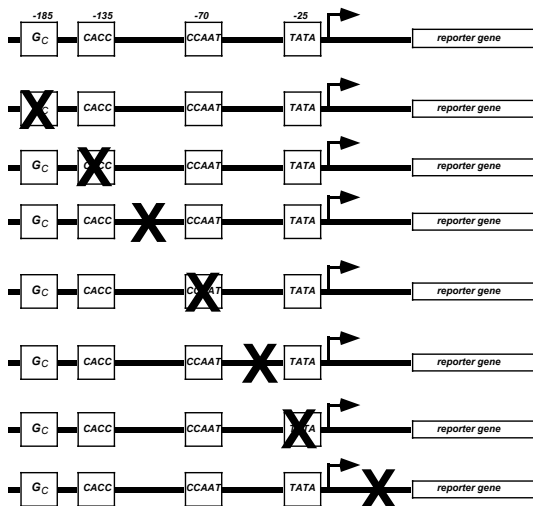
the advantages of nested deletions

- fast
- defines workable DNA/promoter fragment
- can survey a large region

the shortcomings of nested deletions

- only boundaries are defined
- important elements are not identified

c. Linker scanning



the advantages of linker scanning

- defines multiple important elements within promoter fragment
- spatial relationship of flanking elements preserved

the shortcomings of linker scanning

- labor intensive (although most worthwhile data comes from hard work)
- may miss elements due to redundant elements in the promoter

d. DNA binding assays (DNAase I footprinting, EMSA)

after boundaries are defined functionally, DNAaseI footprint with nuclear extracts to look for sites bound by DNA binding proteins in the extract

- look for known binding sites
- mutate sequences protected in footprint (complementary to linker scanning)
the advantages of footprinting
 - directly and immediately identifies factor binding sites
 - easy to look for cell type specific elements (use extracts from different cell types or tissues)
 - can direct the design of promoter mutations

the short comings of footprinting

- not as thorough as linker scanning, may miss sites bound by labile factors.

- The footprint is only as good as the extract.
- May not see factors that bind transiently, or factors that are regulated temporally. Linker scanning might identify these types of sites first.

- EMSA is an alternative assay. It is more sensitive than footprinting but it will now show which sequences are recognized by the DNA binding protein.
-

Trans-acting Factors

One of the main points for this section is that transcription factors are **modular**

Connection to cis-acting elements: Cis-acting elements are comprised of elements that function as binding sites for transcription regulators - there is high flexibility in the arrangement. Likewise, trans-acting factors are also flexible.

In fact, transcription factors give "character" to the promoter - its expression pattern spatially, temporally, its responsive to the "health" of the cell, signals from outside the cell, etc. There are so very many different kinds of transcription factors (est. ~2000-3000) that we cannot even begin to talk about each in this course. The review assigned for this topic (*Brivanlou and Darnell. 2002. Science 295:813*) proposes a new, interesting classification scheme for transcription factors. The classification scheme is based on what these factors do / how they work, rather than their protein structure.

It is instructive to review two of the very first studied eukaryotic, Pol II yeast transcription factors:

GCN4 and Gal4

Gal4: (Brent and Ptashne in 1985)

GCN4: (Hope and Struhl, 1985)

1. GCN4 and regulation of HIS3 gene

- need this factor for coordinate induction of 30-50 genes encoding amino acid biosynthetic enzymes

- eg. HIS3 gene
 - i. induced 10X in aa starvation (trp/arg/his/lys/leu/ile-val/gln/met)
 - ii. induction requires
 - a. cis-acting element -->UAS_{GCN} at -90 bp from start site
 - b. trans-acting factors --> GCN4, a 48 kD protein, binds as dimer

To analyze the protein, make a set of N-terminal deletion mutants:

2 assays:

1. DNA binding - EMSA technique
result: DNA binding with C60 (aa245-281)
2. activation of gene expression *in vivo* lost with C163

look at this by making a yeast GCN4⁻ strain (HIS3 gene not expressed, require histidine in the media), introduce GCN4, look at ability of yeast to grow without histidine in media, indicative of increased HIS3 gene expression

results: 5' boundary between **aa107, aa119**

How to determine the 3' boundary by nested deletion? DNA binding activity will be lost.

method: I. Fuse to heterologous DNA binding domain (lexA from E.coli)
II. Change UAS_{GCN} to UAS_{lex}

Side note: lexA binds and represses genes in SOS response until DNA damage induces a protease. LexA DBD from aa1-81

results: 3' nested deletion of the fusion protein shows: aa12-125 - still robust
aa12-120 - weak
GCN4 activation domain between aa107-125!!!! small!!! 19 aa!

conclusions:

GCN4 has 2 separable activities

- 1. DNA binding - C terminal 60 amino acids**
- 2. Transcription activation - aa107-125**
are the amino acids in between the domains a flexible linker?
- 3. activation mediated by protein•protein contact - not by contact with the DNA binding domain**
- 4. DNA binding domain serves to anchor protein to the promoter**
- 5. swivel between domains allows orientation independence**
- 6. looping allows distance-independence**

2nd example:

Galactose catabolism in yeast - many enzymes needed.

eg. GAL1 => galactokinase

gene looks like:

Galacto-kinase activity		<u>glycerol</u>	<u>Gal</u>	<u>Glu</u>
	WT	50	810	7
	Gal80	50	810	50
	Gal4	50	50	50

conclusions:

1. GAL1 expression activated by galactose, repressed by glucose
2. GAL80 necessary for repression
3. GAL4 necessary for activation and repression

- GAL4 binds UAS_{gal} => positive regulator 4X17bp elements
- Ma/Ptashne map functional domains by deletion analysis
- artificial reporter gene in yeast $UAS_{GAL4}=>LacZ$

Table I conclusion: activity lost (aa792), regained (aa238), lost again (aa196)

Table II conclusion: activation domain aa 768-881

Table III conclusion: 2nd activation domain activity is just as good (if 147-238 is added)
both activation domains together, roughly 2X

Summary of Structure:

1. again, see modular design
2. 2 activation domains - roughly independent

The modular design of GAL4 and the well-studied HIS3 gene have been exploited for use in a relatively new and powerful new technique called the yeast two hybrid assay:

Many years since the analysis of GCN4 and Gal4, structure/Function analysis of myriad transcription factors has been performed - the majority are modular in design

can classify

A. DNA binding domains - fold independently, can crystallize small proteins to get structures with and without DNA. By now, the structure of many DBD domains are known

B. transcription activation domains

DNA binding domains

dimerization is a major theme among DNA binding domains

a. dimers allow mixing/matching - a good way to modulate DNA specificity and affinity

a good way to create many regulatory/functional properties

b. how to make many different activities from a relatively modest # of regulators

c. potential for dominant negatives

- popular point of genetic therapy
- naturally occurring examples
 - myoD/Id - bHLH/b_pHLH (called E box factors)
 - steroid receptors - Zn finger DBD

dimer binding elements tend to be palindromic

eg. E box: CANNTG

steroid hormones: GGTACAnnnTGGTCT

a monomer binding site to compare:

eg. LEF-1: TCTTTGWW

The diversity provided by dimerization between different members of a factor family offers enormous scope for regulation. Different combinations have distinct regulatory and functional properties.

d. DNA binding affinity

bHLH, bZIP families: Jun

low affinity: jun₂ dimers, can bind the site TGAGTCA

hi affinity: jun•fos heterodimers also bind this site - this binding is more stable

e. DNA site specificity

eg. jun + fos -> TGAGTCA
CREB₂ -> TGACGTCA

eg. steroid receptors (members bind either as homo- or hetero- dimers to HRE's)

GRE: GGTACA nnn TGGTCT

TRE: AGGGTCA nnnn AGGTCA - here n=4, n=3 (VDRE), n=5 (RARE)

orphan receptors, classified on the basis of the spacing of their half sites

heterodimer formation: RXR -> VDR, TR, RAR each bind to slightly different sites

f. Dominant negatives

One example: bHLH proteins in mammals and Drosophila

E box proteins are members of the bHLH type of DNA binding protein. They are important for developmental control of gene expression in a number of different systems: muscle differentiation, peripheral nervous system, and the immune system

mammals

myoD family

Drosophila

AS-C (achaete-scute complex)

da (daughterless)

Id

h (hairy)

emc (extramacrochaetae)

myo D, myogenin, E12, E47, AS-C, da -> positive role to promote differentiation.

Loss of their activity -> lose organ development

h, emc, Id -> negative regulators. Loss of their activity -> increase # of size/speed of organ development

Id proteins are negative regulators because they are dominant negatives: thus in the developmental program, Id is present in myoblasts to prevent the actions of myoD. When myoblasts are induced to differentiate into myotubes, they turn off the expression of Id, allowing myoD to activate target genes.

<u>myoblasts</u>	->	<u>myotubes</u>
myoD ON		myoD ON
Id ON		Id OFF

g. Heterodimer or protein•protein interaction on the DNA

what about the DBD's that do not form dimers?

WHTH (eg. the Winged Helix-turn-Helix, ets protein), DNA binding can be modified by flanking factors

The combinatorial nature of transcription

- different combinations of binding sites generate different transcriptional patterns
- different combinations of dimer partners can have different functional activities
- therefore understanding how the expression of a particular gene is controlled in a given cell presents a daunting prospect. Besides identifying the essential DNA sequence elements, it entails identifying all the proteins that have the potential to bind these sites, measuring their abundance and distribution in the cell and detailing their interactions and consequences of those interactions

Activation domains

more difficult to study
limited biophysical data that are available suggest poor structure in solution
less well-defined - hard to determine consensus aa sequences so they are be
classified by primary aa sequence, but mutational analysis would suggest that other
residues in the domain are important
acid blobs
pro-rich
Q-rich
ser/thr-rich

does this mean they are induced structures? will need to identify their direct targets.
Some targets have been identified, such as the GTFs and the Mediator. But others are
referred to as co-activators or components of chromatin modification/remodeling
complexes and we will discuss them in the next topic.