Chapter 28  
Regulation of Gene Expression

28.0 Intro
4000 genes bacterial genome 25,000 in human  
only a fraction is expressed at any one time  
some gene products needed in large amounts, others, only a few per cell  
enzymes needed for a given pathway may be needed for only a little while

Cellular conc. of a protein determined by a balance between at least 7 process
  1. Synthesis of primary RNA transcript  
  2. post-transcriptional processing of mRNA  
  3. mRNA degradation  
  4. Protein synthesis  
  5. Post-translational modification of protein  
  6. Proteins targeting and transport  
  7. Protein degradation

Figure 28-1

While control can, and, is expressed at all 7 levels
this chapter deals primarily with initiation of transcription  
is most common and best understood process right now  
also, since it is right at the beginning, is most effective, so is most common

28.1 Principles of Gene Regulation
   *Housekeeping Genes or constitutive gene* genes expressed at a more or less constant level because needed constantly

   *Regulated Gene Expression* levels of gene product rise and fall in response to molecular signals

   *Inducible Gene* - gene products that increase in concentration due to a particular signal

   Process called *induction*

   *Repressible Genes* - gene products that decrease in concentration due to a particular signal

   process called *repression*

Much of transcriptional control is mediated at the RNA polymerase/DNA binding step  
Let’s start there
A. RNA Polymerase Binds to DNA at Promoters

Saw in chapter 26 RNA polymerase regions binds at sites called promoters
- Generally near where RNA synthesis will begin
- Regulation will involve modulating this interaction

Brief review figure 28-2
- Sequences in promoter region vary widely
- In general closer to consensus, more often transcribed
- Further from consensus less transcribed
- May effect by factor of 1000

Constitutive genes not expressed at same levels due to this difference
- Regulated gene involved this + additional modulation by regulatory gene products
  - Often either enhance or interfere with binding to promoter regions

Eukaryotic promoter regions more variable
- 3 eukaryotic polymerases need an array of additional factors to bind to promoter sites

B. Transcription Initiation is regulated by proteins that bind at or near promoters

3 types of proteins regulate transcription
- Specificity factors - alter specificity of RNA polymerase for a promoter (or set of promoters)
- Repressors - impede access of RNA polymerase to promoter
- Activators - enhance RNA-promoter interactions

Specificity Factors

Already talked about specificity factors in changer 26, but didn’t call them specificity factors at that time. Can you guess what they were?

- σ factors
  - $\sigma^{70}$ (70,000 MW) most common - recognizes most promoters
  - 6 other specificity factors
  - One is $\sigma^{32}$ (32,000 MW) promoters for genes related to heat shock response
    - Different consensus

Figure 28-3
Allows for the coordinated expression of several protein products at once
Several equivalent proteins in Eukaryotes
   In particular TBP TATA-binding proteins

Repressors Figure 28-4 A&B
   Bind to specific DNA sites called operators
      Generally near promoter
      RNA polymerase either can’t bind, or it binds, but can’t get to where it should be
      Referred to as negative regulation

   Binding of repressor can be regulated by other binding events
      Either other proteins or small molecules
      Called effectors
      Binds to protein to make conformational change
      Change either increases or decreases binding of repressor
      In turn decreases or increases transcription
      In some cases complete dissociation of repressor from DNA
      Another case binding of effector makes repressor bind

Eukaryotic cells similar, but repressor may be more distant

Activators Figure 28-4 C&D
   Positive regulation
   Their binding enhances binding of Polymerase to promoter
   Activator sites usually adjacent to promoters
      Some times no interaction without promoter

Eukaryotes
   Enhancers (Eukaryotic equivalent)
      Can be 1000's of bp from promoter

   Sometimes enhancer normally bound helping gene express
      And gets dissociated by a molecular signal
   Other time is not bound until molecular signal make conformational change

   Signal can increase or decrease transcription
   Positive regulation common in Eukaryotes
      Also more complicated
      Reason can be 1000's of bp away is that intervening DNA gets looped out Figure 28-5 by proteins called architectural regulators
C. Most Prokaryotic Gene are clustered and regulated in **operons**
   - simple mech for coordinated regulation
   - whole set of gene clustered on chromosome and transcribed in 1 piece
     - works well in prokaryotic because polycistronic
       - (several genes on one piece of DNA)
   - Gene cluster + promoter + additional sequences that function together
     - called **an operon** Figure 28-6
   - Common size 2-6 genes
   - Some up to 20 or more

   Term operon first introduced 1960 by Jacob & Monod

   Described the lac operon
   - Genes that have to do with lactose metabolism

D. The Lac Operon - an example of negative regulation
   - Figure 28-7
     - Need permease (Y gene) to get lactose into cell
     - Need galactosidase (Z gene) to split into monosaccharides
       - also includes a thiogalactoside transacetylase (A gene)
     - Modifies toxic galactosides for removal?

   Each gene includes a ribosome binding site for independent translation
   - (not shown in figure)

   Figure 28-8
   - In absence of lactose operon is repressed
     - Repressed by binding of protein called the lac repressor (the I gene)
     - Is a tetramer of identical monomers
     - Is coded for on by a different gene with a different promoter (P_I)
     - That happens to be just upstream of lac operon
   - Binds at three different sites on gene
     - O_1 tightest binding
       - Right at RNA polymerase start
         - (See figure 28-11)
     - Two other binding sites
       - O_2 inside Z gene
       - O_3 inside I gene
     - (Note: 1 dimer binds at O_1, a second at O_2 or O_3, so is tetramer overall)
   - To repress the inhibition must bind to O_1 and either O_2 or O_3, looping out intervening DNA
Control not absolute
   Down about 1,000 when repressor is functioning
   If eliminate O₂ and O₃ so just have O₁ down about 100

   So even when repressed some low basal level of expression

This basal level is needed for induction

Induction
   The few permeases let lactose into cell and galactosidase converts
to allolactose (an intermediate before gets to monosaccharides?)

   Allolactose binds to repressor
      Conformational change
      Released from DNA
      Conc of lac proteins increases by 1000

Several substance can also bind to repressor and act as inducers
You have probably used IPTG in lab
   Isopropylthiogalactosidaase (structure right column 1160)
      Cannot be metabolized so turns on gene

Actually more complicated than shown here
There is an additional activating factor as well
will discuss multiple layers of control later in chapter, for now just getting
the basics down

Now many polycistronic operons identified in bacteria and a few in lower
Eukaryotes

Most eukaryotes are monocistronic so each gene controlled separately

E. Regulatory Proteins have discrete DNA-binding domains

Regulatory proteins generally bind to specific DNA sequences

Affinity 10⁴ to 10⁶ higher that random DNA
Usually have discrete DNA binding domain
   Usually one of a few recognizable DNA binding structural motifs
Must be able to recognize different DNA sequences
   Surprisingly ? don’t need to open up DNA
   Can get it directly from Major groove or minor groove
   Figure 28-9
   Do this mostly with H bonds
Most often use Asn, Gln, Glu, Lys or Arg
Gln & Asn form 2 bonds with N\textsuperscript{6} and N-7 of A and no others
Arg can make 2 bond with N-7 and O\textsuperscript{6} of G and no others
(See figure 28-10)
But CH\textsubscript{3} of Thymine used to distinguish from C

Several other ways. No exact AA to base code
Can also do via minor grove but not as easy

Only a small piece of protein needed to interact with DNA
DNA binding domains tend to be small (60-90 residues)
Actual amount of protein actually touching DNA is even smaller

Binding domains near minimum size for stable hydrophobic in
hydrophilic out structure. Built very carefully or made as a bulge on
a bigger protein

DNA binding sites usually inverted repeats or palindromes
Easy to use protein dimer to bind to both sites as once
Lac repressor unusual with tetramer structure
Two dimers at 1 O\textsubscript{1} site
Other two dimers at second site (O\textsubscript{2} or O\textsubscript{3})
(Figure 28-8B)
Each dimer site includes contacts with 17 of 22 bases
Shown figure 28-11
Binding at O\textsubscript{1} has a K\textsubscript{dis} of 10\textsuperscript{-10} M
So very specific

Several DNA binding domains are recognized
Will focus on 3 most common in DNA regulatory proteins
Helix-turn-helix
Zinc finger
Homeodomain some eukaryotes

Helix-turn-helix
(Figure 28-11)
Seen in many prokaryotes and similar seen in some
eukaryotes
7-9 residues of helix
A beta turn
7-9 residues of helix
Total of about 20 resides

Structure not self stable
Bulge out of a larger stable protein
One helix called recognition helix because it is placed in major groove of DNA has DNA interactions

This is motif used in lac repressor

**Zinc Finger Figure 28-12**

- Used in many eukaryotes
- 30 residues
  - 4 are cys
  - Or
  - 2 cys and 2 his
- Coordinate a single Zn$^{2+}$
- Zn$^{2+}$ not part of DNA interaction
- But is the core that holds the motif together

DNA interaction with a single finger usually weak

Need several finger for better binding

Mouse regulatory protein Zif268 uses 3 Zinc fingers in a single polypeptide to bind DNA

Frog DNA binding protein uses 37!

A wide variety of DNA-protein binding interactions are used

Also use in RNA binding

**Homeodomain Figure 28-13**

- Used often in eukaryotic developmental regulators
- 60 AA
- Called homeodomain because discovered in homeotic genes - the genes that regulate development of body pattern

- Highly conserved and observed in many organisms

- Similar to helix-turn-helix motif

- Gene coding for domain is called the homeobox

**F. Regulatory Proteins also have protein-protein interaction domains**

- Regulatory proteins need to have protein/protein interactions
- Bind to themselves to make dimers
- Bind to RNA polymerase
- Bind to other regulatory proteins
- Bind to transcription factors
Again a few common motifs are seen often
  Leucine zipper
  Basic Helix-loop-helix

Leucine zipper (figure 28-14)
  Amphipathic α helix hydrophobic A’s run on one side
  See a leu every 7th residue (that where gets name
  Hydrophobic surface used to hold a dimer of proteins together

  Originally thought that leu’s interdigitated like a zipper
  Now know that side by side in a coiled coil

  Protein with leu zippers often have separate DNA binding domain
  with lots of Arg and Lys

  Note: figure is a little misleading because it almost looks like a
  continuous Helix-turn-helix from one protein. Actually a Helixes
  from 2 different proteins

  Found in many eukaryotes and a few prokaryotes

Basic Helix-loop-helix (figure 28-15)
  Used in eukaryotes control of gene expression in multicelluar?
  Conserved region about 50 AA that does both DNA binding and
  dimerization

  1 helix is DNA binding - rich in basic AA’s
  Then a variable length loop
  2nd helix is dimer interface

  Structure distinctly different from helix-turn helix where one helix
  and turn did DNA binding and second helix was for structural
  support

Protein-Protein Interactions in Eukaryotic Regulatory proteins
  In Eukaryotes
    Most genes regulated by activation
    Most genes moncistronic
  If needed a different activator for each gene would need 1000's of
  activators
  Yet in yeast only about 300 transcription factors(mostly activators)
  Most transcription factors activate multiple genes
  Most genes regulated by multiple transcription factors
  So control is achieved by utilizing different combinations of a
  limited number of transcription factors
This is called **combinatorial control**
Several families of eukaryotic transcription factors defined based on mix and match of the above (and a few other) structural motifs

Several families of transcription factors based on close structural similarities
Part of combinatorial control is based on mixing and matching members
  So get both homodimers and heterodimers

So a family of 4 different leucine zipper binding proteins could make up to 10 different dimeric species
  AA, AB, AC, AD, BB, BC, BD, CC, CD, DD
Each dimer can have distinctly different binding properties

So get a wide range of diversity with just a few proteins

Also need to interact with RNA polymerase other regulatory proteins or both

At least 3 additional protein/protein interaction domains have been recognized (primarily in eukaryotes)
  Glutamine rich
  Proline rich
  Acidic domains

**28.2 Regulation of Gene Expression in Prokaryotes**
Prokaryotes simpler so will do first
presenting a few well understood systems as overview, not exhaustive list also similar to things will see in Eukaryotes

**A. The lac operon (continued)**
  Last saw was a single repressor on/of type control
  Too simple
  Want other controls as well
    For instance glucose is preferred E source
    So want to shut down lac operon, if glucose is present regardless of whether lactose is also present

  A second control mech called *catabolite repression*
    If glucose present
      Shuts down genes for lactose, arabinose and others
Effect mediated by cAMP and cAMP receptor protein CRP
CRP also called CAP catabolite activator protein
Figure 28-16 & 28-17

CRP/CAP 28-16 & 28-17
Homodimer of 22,000MW proteins
Binds both DNA and cAMP
  Binding to DNA ↑ in presence of cAMP
  Binding done by helix-turn-helix motif
  Note shown in figure
Binds to RNA polymerase and DNA
Used to make RNA polymerase bind better to weak promoters

Glucose absent (cAMP ↑, Binds to CRP)
  CRP binds to site near lac promoter (see fig 28-17)
  Increases RNA transcription 50X
  Therefor glu↓ lac↑ so is positive regulator

Two effectors act in concert
  CRP has no effect one way or other if lac repressor is in place
  However if lac repressor released then weak lac promoter doesn't get much going unless CRP is bound

So need both lac to be present and Glu to be absent

How does cAMP play into this?
  CRP has a cAMP binding site
  Bind of cAMP increases binding of CRP to DNA

  When [Glucose] high
    Synthesis of cAMP is low
    AND cAMP is transported outside of cell
  Net [Glucose]↑ , [cAMP]↓ binding of CRP ↓ transcription of lac↓
  [Glucose]↓ , [cAMP]↑ binding of CRP ↑ transcription of lac↑

CRP and cAMP involved in coordinated regulation of many operons
  Lactose , arabinose and others

Network of operons regulated by a common regulator called a regulon
  Can be used for coordinated expressing of 100's of genes

Will look at another regulon, the SOS system later in chapter
B. Transcription attenuation (Common in AA biocynthetic pathways)

mech used for many genes using in AA biosynthesis

E coli can synthesize all 20 AA’s
enzyme for synthesis of a given AA usually clustered into an operon
operon expressed only when external supplies of that AA are inadequate

tryptophan operon is a good example (figure 28-18)
5 proteins need to make tryptophan
Some proteins do more than 1 reaction

mRNA for this transcript had half-life of about 3 min

Has a normal repressor
Trp repressor is a dimer
When trp present, bind to repressor, repressor binds to operator
Operator site overlaps promoter site so when bound can
start transcription complex

Simple on/off not enough

Figure 28-19
Can see additional fine tuning control mech
Mech relies on close coupling between translation and transcription
in bacterial cell

Notice that between promoter and 1st trp gene is a leader sequence
leader sequence contains an AUG so has sequence for a short protein
before real proteins

162 nucleotide leader sequence
essentially a small peptide
Complete with a start, stop and, Most importantly,
The usual hairpin.loop UUU sequence used as a termination
and release. (Back in chapter 26 typical for rho independent
termination) RNA structure usually used to terminate
transcription of DNA into RNA

Also built in are a couple of other hairpins 1:2, 2:3 and the
termination hairpin 3:4

If trp repressor allows transcription of trp operon, it starts and then
ends right here after only 139 nucleotides (45 resides) read off and
before the message for any real protein has be transcribed!!
Hence name of control mech, attenuation

How to release attenuation?

Have mentioned before that ribosomes attach to mRNA even before it is off of DNA, and translation and transcription can be almost simultaneous in bacteria

As mRNA for this leader is being transcribed, it, in turn, is being translated

The first peptide has 2 trp’s in its sequence

When it gets translated if TRP present, they get incorporated and every thing goes as stated

However if TRP absent (because cell really need TRP to be synthesized)
The ribosome stalls at this point

When the ribosome stalls, it stalls on top of the region 1
This makes region 2 form a hairpin with region 3

This keeps 3 from making the hairpin with 4 that signals to end transcription
So RNA polymerase carries on with the rest of the message!!

Many other AA synthetic operons use the same kind of attenuation mech
Pretty neat because don’t need any other proteins and is sensitive to the AA

Leader for PHE attenuation is 15 residues, and 7 are phe
leader for leu has 4 leus
leader for his has 7 his

in His operon attenuation is the only control mech!
C. Induction of the SOS Response

Extensive DNA damage in bacteria triggers induction of many distant genes used in DNA repair (see figure 28-20)

Called SOS response

Coordinated control of several distinct genes

Key players

Rec A protein

Should remember from chapter 25 page 1041. Forms a protein filament around single stranded DNA Figure 25-32

LexA repressor

LexA repressor

22,700 MW

Inhibits transcription of all SOS genes

But not simple repressor

Repressor activity inactivated by its OWN -self cleavage into two roughly equal peptides

At normal pH this cleavage requires RecA protein

But RecA not a protease

Its interaction allows LexA to cleave itself

RecA must be bound to single stranded DNA before will bind to LexA

This is link to SOS

Only when cellular DNA is severely damaged will enough gaps exist in DNA so RecA will bind to single stranded gaps. Once it binds, it activates the LexA to cleave itself, once LexA cleaves itself, the repression of repressed genes is removed so start copying SOS repair genes

Some bacteriophages have adapted this system for their use

When Cell has damage, RecA binds to single strand DNA

Starts helps LexA, and some repressors that have kept bacteriophage genes suppressed both self cleave. Bacteriophage now replicates and gets a chance to abandon ship as cell dies from the bacteriophage lysis
D. Coordinated Synthesis of Ribosomal Proteins and rRNA

if bacteria need more proteins synthesized, will increase number of ribosomes
a general correlation between # of ribosomes and cellular growth rate

Need to coordinate synthesis of ribosomal proteins and RNA

A distinctly different control mech, works via at translation level rather than transcription.

52 genes for ribosomal proteins
20 operons
  Each operon between 1 and 11 proteins
  Also in some operons are:
    DNA primase
    RNA polymerase
    Protein synthesis elongation factors

  Thinks this helps couple replication, transcription and translation

Translation feedback control of r-proteins (ribosomal proteins)
  Ie. Binds to mRNA to prevent ribosomes from making proteins
  So Binding to RNA not DNA
  Each operon in system also codes for a translational repressor
  Binds to mRNA from the operator to keep from being translated!
  See figure 28-21
  The repressor also binds to rRNA with higher affinity

So will only repress mRNA of proteins if [protein]>[rRNA]
  So as protein goes into excess it represses itself!

Binding site for translational repressor is near start of mRNA
Unlike transcription, each protein in an mRNA is usually translated independently

  Only in these operons is translation linked, so if you stop translation of the first gene all others are stopped

  Why this happens is not understood
  May be tied to 3D structural fold in mRNA

There is also a transcriptional control of ribosomal proteins
More transcription as growth rate increases
Mech not understood
Just saw Protein tied to level of rRNA
How is rRNA controlled?

Synthesis of 7 different rRNA operons controlled by cellular levels of nutrients, in particular AA’s

Control mech is called *stringent response* (figure 28-22)

When run out of AA’s, ribosomes stall and halt on mRNA
Uncharged AA come in and binds at A site
When this happens, a factor called *stringent factor* also binds to ribosome (stringent factor is actually RelA Protein)
When stringent factor binds it does the reaction:
\[
GTP + ATP \rightarrow ppGpp + AMP
\]
Step 1: GTP (pppG) + ATP \rightarrow ppGpp + AMP
Step 2: ppGpp \rightarrow ppGpp

The ppGpp is signal that slows rRNA synthesis, in part, by binding to RNA polymerase

Have now seen cAMP and ppGpp as modified nucleotides
Use as second cellular signals
In this case for starvation
Eukaryotic cells also use similar modified nucleotides as signals
More will probably be found

E. Function of some mRNA’s is regulated by small RNAs in Cis or Trans
RNA control of gene regulation is just now becoming understood
(This section was not present in 4th edition)

Functions of mRNA can be controlled by r proteins (just saw above)
Or by RNA

Controlling RNA can be within the mRNA itself or an entirely separate RNA
If RNA is within the mRNA, called acting “in cis”
When controlling RNA is separate from mRNA called acting “in trans”

Example 1: regulation of mRNA for RNA polymerase sigma factor (rpoS)\[\sigma^S\] (remember what a sigma factor is?)
Used when cell under stress from lack of nutrients
And needs to enter stationary phase
\[\sigma^S\] used to express large number of stress response genes
\[\sigma^S\] usually expressed at low levels
But not translated because hairpin forms that inhibits
ribosome binding (figure 28-23)
Under stress conditions one or both of two small special function RNA's are induced
- DsrA (downstream region A)
- RprA (Rpos regulator RNA A)
Either can bind with ½ of hairpin
  - Disrupts hairpin
  - Allows ribosome to bind
Other samples exist
- All rely on 'small' RNA's
  - <300 nucleotides
- Also require protein Hfq
  - RNA chaperone that helps make RNA-RNA pairing
Not very common. Probably only a few dozen genes in a bacteria use this system
More common in Eukaryotes

Example 2: in cis riboswitches
Box 26-3 figure 28-24
Riboswitchs - aptamers of RNA molecule
- Aptamer a RNA that binds a small molecule
- Aptamer built into 5' end of mRNA
If binds to its signal molecule
  - Can make structure to encourage termination of translation
  - Can make structure to discourage ribosome binding
Most genes using this mech are gene involved in synthesis or transport of the molecule that binds to RNA aptamer
  - Or if that molecule is present, no need to translate message
Riboswitches have been found for over a dozen ligands
Drugs now being found to bind various switches to turn off key genes in bacteria

F. Some Genes regulated by genetic recombination
used in Salmonella bacteria that live in human gut
- have flagella that use for motility
- flagella made with many copies of protein flagellin
target of mammalian immune system

Bacteria switches between FljB abd FljC every 1000 generations through process called phase variation
  - A way to avoid immune response?
Controlled by site specific inversion of promoter sequence
Performed by site specific recombination done by recombinase called Hin

In one orientation promoter turns on fljB and fljA
    The B is the flagellar protein
    The A is a repressor to keep fljC turned off

In other orientation does not express B or A
    Repression is lost and fljC starts up

Not a unique system
recombination systems have been found in other prokaryotes as well as eukaryotes

28.2 Regulation of Gene Expression in Eukaryotes
eukaryotes also use transcriptional control, but will have several differences
‘Transcriptional ground state’ inherent activity of transcriptional activity in absence of regulatory sequences

In bacteria RNA polymerase generally can access all promoters so can initiate transcription unless specifically turned off.
    Called a non-restrictive ground state

In eukaryotes promoters generally turned off, and you need a promoter to turn on
    Called a restrictive ground state

Why and how are Eukaryotes different from prokaryotes
1. Access to gene is restricted by chromatin structure
    Several changes must occur in chromatin structure before a gene can be transcribed
2. Both + and - control elements in Eukaryotes, but + is dominant
3. Eukaryotes use large complex regulatory proteins
4. Translation and transcription separated in time and space
A. Chromatin Structure

Transcriptionally active DNA structurally different than inactive Chromatin

Transcription strongly repressed when DNA condensed in chromatin nothing equivalent in prokaryotes

While a chromosome my look dispersed and amorphous where are actually some distinct forms of chromatin

Heterochromatin - more condensed - transcriptionally inactive

Usually about 10% of chromosome

Euchromatin - less condensed -some but not all is transcriptionally active

Transcriptionally active

More open structure

Nucleosome have a particular composition and types of modification

Deficient in H1

Enriched inH3.3 & H2AZ

Methylation

Acetylation

DNA is eukariots often methylated on C of CpG

Undermethylated when transcriptionally active

Overall thought, physical changes must occur in DNA, histones and chromatin before it can become transcriptionally active

5 known families of enzyme complexes reposition or displace nucleosomes while hydrolyzing ATP

3 particularly important in transcriptional activation Table 28-2

SWI/SNF

Found in all eukaryotes

At least 6 core polypeptides

Remodel chromatin so nucelosomes are irregularly spaced

Stimulate transcription factor binding

NURF

Member of ISW1 family

Remodels Chromatin

Complementary and overlap activity of AWI/SNF

SWR1

Enriches histones with H3.3 and H2AZ

Histones also get deficient in H1 as become more transcriptionally active
Other changes to histones in transcriptionally active chromatin

Core histones (H2A, H2B, H3, H4)
- Methylated at lys or Arg
- Phosphorylated at Ser or Thr
- Ubiquitinylated, acetylated, sumoylated
  - SUMO (small ubiquitin-like Modifier)
  - A protein that gets attached like ubiquitin

Remember structure of histones? Figure 24-26
- Well structured core and unstructured amino termini?
- Modification occur at specific residues in unstructured amino terminii
- Patterns of modification may be a ‘histone code’ for protein recognition

Acetylation and methylation are prominent in active chromatin
- First methylated by specific methylases at specific lys
  - The bind HAT’s
    - (Histone acetyltransferases)
  - And acetylate particular Lys
    - When first synthesized in cytosol
    - Type B HATS acetylase
      - Then transported into nucleus
      - Assembled into nucleosome with help of other proteins
      - Bind to DNA to make chromatin with help of
        - Histone chaperones CAF1 & NAP1

When nucleosome activated for transcription
- Further acetylation by
  - Type A (nuclear) HAT’s
  - Seems to reduce affinity for DNA
  - May also have regulatory protein-protein interactions

When no longer actively transcribed
- Deacetylated using histone deacetylases (HDAC)
- Also lys\(^9\) of H3 methylated
C. Many Eukaryotic Promoters are + regulators
most eukaryotic RNA polymerases have no affinity for promoters
most need several activators to get things started

Why
Why not repressors?
If chromatin blocks access to gene, repressor redundant

Multiple activators
In large chromosomes more chance that a given regulatory
sequence will occur randomly
With multiple sites to promote, less chance of accidental random
initiation

Why promoters?
With 25,000 genes would need 25,000 repressors
If everybody repressed, then only need a few activators to activate
sets of genes as needed

With promoters can activate genes on several chromosomes
simultaneously

In spite of above logic, don’t be fooled, there are repressors

D. DNA Binding transactivators and coactivators help assemble general
transcription factors

Chapter 16 learned that mRNA synthesized by RNA polymerase II (Pol II)
Common features of Pol II promoters were:
TATA box about -30
Inr box about 0 (initiator)
Figure 26-8
And other regulatory sequences

Now about those other regulatory sequences
Usually called enhancers in higher eukaryotes
Called upstream activator sequences (UAS) in yeast
In Yeast almost always upstream
And almost always within a couple of hundred bp

In other eukaryotes may be several 100 or even 1000 bp
upstream!
May also be downstream
May also be in gene itself!
Generally bind regulatory protein and that increases transcription of any promoter in area, upstream or downstream.

Usually very complicated because an average of ~6 positive regulators are used in any given interaction.

Five Classes of Proteins required for successful binding of RNA pol II

**Figure 28-28**
- Transcriptional activators
  - Bind to enhancers or UAS to facilitate transcription
- Architectural regulators
  - Facilitate DNA looping
- Chromatin modification or remodeling proteins
  - (Described earlier)
- Coactivators
  - Go between - does not bind to DNA
  - Bridge between (Basal transcription factors and Pol II) and transcriptional activators
- Basal transcription factors
  - Required by most Pol II promoters

Details

**Transcriptional activators**
- Binds to DNA sequence called enhancer
- Some used for hundreds of promoters
- Some for only a few
- Many sensitive to small molecule binding for activation or deactivation
- Enhancer region may be distant from TATA box
- Typically 6 or more enhancers bound by 6 or more transcriptional activators required to provide combinatorial control

**Architectural Regulators**
- Since many enhancers are far from TATA box need to loop out intervening DNA
- These are protein that bind DNA with limited specificity
- Abundant in Chromatin
- Most prominent - HMG proteins
  - High Mobility Group - Runs fast on gel
Coactivator complexes
Intermediates between Pol II complex and transcription activators
One complex called Mediator
- 20 or more highly conserved peptides
- 4 more subunits can inhibit transcription
- Binds tightly to CTD (carboxy-terminal domain) of Pol II
- Required for both basal and regulated transcription
- Also stimulates phosphorylation of CTD by TFIIH

Coactivator complexes function at or near promoters TATA box

TATA-Binding Protein
Review from chapter 26
- Binding of Pol II typically starts with TATA binding protein (TBP) binding at TATA sequence to make preinitiation complex
- TBP often delivers in complex with ~ 15 other subunits
- Binding of TBP and Pol II not enough for transcription
- Need lots of other stuff to fall into place

Now you know the players lets look at how it works

Choreography of transcriptional event Fig 28-29
Exact order may vary, but this is a nice starting point
1. some activators have strong enough binding can find site even when covered in chromatin
2. binding of one activator helps others to now bind
3. activators now interact with HAT’s or complexes like SWI/SNF
   Remodel surrounding chromatin
4. Activators now interact with Mediator complex
5. Mediator acts as a scaffold to assemble TBP or TFIID, then TFIIB
6. Other components of preinitiation complex (PIC) including Pol II come together
Details are complex and vary

Reversible transcriptional activation
- Some proteins to repress binding of RNA pol II do exist, but are rare
- Some activators have multiple conformations can act + or -
  Seen in some steroid hormones
- When steroid binds, activator activates
- When steroid absent, receptor prevent formation of
preinitiation complex

In some cases repression involves restoring histones and chromatin to inactive state

In some cases repressors bind to mediator to block transcription

E. Example gene - Galactose metabolism in yeast both + and - control
well studied system in Yeast
Figure 28-30 Table 28-3
genes required for important and metabolism spread throughout several chromosomes
Each GAL gene transcribed separately, no operon structure
all gal genes have similar promoters
all have TATA box, Inr sequences and an upstream activator, UAS₉
UAS₉ recognized by DNA-binding transcription activator Gal4p
regulation includes interplay between:
Gal4p, Gal80p and Gal3p

Gal 80p forms complex with Gal4p to prevent functioning as an activator (Still binds to DNA?)

Galactose (when present) binds to Gal3p,
This complex binds to Gal4p/Gal80p complex, and releases 80p
Gal4p now acts as activator for gal promoter
As Gal gene products build up Gal3p may be replaced with Gal1p (a galactose kinase) that sustains activation of Gal genes

Other protein complexes involved
SAGA complex - histone acetylation
SWI/SNF complex - chromatin remodeling
Flavor of how complicated figure 28-30
Most of this works through the Gal4p protein
Also has a catabolite repression system as in e coli, so whole thing is suppressed if glucose is present includes even more proteins not shown in above figure
F. Transcription Activators have a modular structure
usually a DNA binding domain
one or more transcriptional activator domains
can have domains for interactions with other regulatory proteins

Interaction between regulatory proteins often mediated by domains
containing leucine zippers or helix-loop helix motifs

Look at 3 mains types of domains used in activation by DNA binding transactivators that come from three proteins Gal4p, Sp1, and CTF1

Figure 28-31a

Gal4p
- Zinc finger near n-terminus of DNA binding domain
  - 6 cys hold 2 Zn²⁺
- Functions as a homodimer (uses coiled coil to hold together)
- Binds to UAS₉ a 17 bp palidromic DNA
- Contains a separate acidic activation domain
  - Can vary sequence of domain a bit, and it will still work
  - But can’t get rid of acidic residues

Sp1
- MW 80000
- DNA binding transcription activator for a large number of genes
- DNA site called a GC box
  - Consensus sequence GGGCGG
  - Usually near TATA box
- DNA binding domain near COOH end of protein
  - Contain 3 zinc fingers
- 2 other domains
  - Both are glutamine rich domain (25% residues GLN)
  - Similar domains seen in many activator proteins

CTF1
- CCAAT-binding transcription factor 1
- Part of a family of transactivators that bind at CCAAT site
- Consensus TGGN₉GCCAA (N is any nucleotide)
- DNA binding domain is basic and probably an α helix
- Not one of our familiar motifs
- Details still being worked out
- Has a proline rich domain (20% pro)
When done right DNA binding domain and protein interaction domains can be swapped between proteins so they are somewhat independent. Interestingly, the 4th edition of this text, said that these kinds of experiments did NOT work!

**Figure 28-31b**

**G. Regulation by intercellular signals**
steroid hormones (and thyroid and retinoid hormones) have additional regulation on Eukaryotic genes

Too hydrophobic to be free in blood
- Travel on specific carrier proteins in blood
- Get to target cell, and can readily pass through PM and get into nucleus
- Bind to specific receptor protein in nucleus
- Hormone-receptor binds to highly specific DNA sequences called HORMONE RESPONSE ELEMENTS HRE’s
- Receptor protein change conformation and interact with additional proteins
- These interaction either enhance or suppress adjacent genes

As shown in **Figure 28-32** Two types of steroid binding nuclear receptors
- Both use HRE’s
  - Type I receptors found in cytoplasm and move to nucleus when bind hormone
  - Type II receptors always in nucleus
- Don’t worry about rest of details hidden in figure legend
- Instead concentrate on HRE’s
- Consensus sequence for HRE’s similar in length and arrangement, but differ in sequence for each hormone
  - See table 28-4 for sequences

Sequences usually 2 six base segments
- Either adjacent or 3 nucleotides apart
- Can be either tandem or palindromic repeat

Hormone receptors - **(Figure 28-33)**
- Highly conserved DNA binding domain - 2 Zn fingers
- Hormone binds as a dimer
  - Each Zn finger binds 6 bp segment

Ability of hormone to act through receptor depends on
- Exact sequence of HRE, relative position to the gene, and # of HRE’s
Ligand binding domain always at COOH end of protein
Each binding domain is unique, no common sequence
   As little as 17% sequence homology
Can vary in size from 25 to 603 AA’s
A single mutation can sometimes completely destroy function

Some hormone receptors use **steroid receptor RNA (SRA)**
   As coactivator
   700 nucleotide RNA part of protein RNA complex
   RNA is required part of complex

H. Regulation can occur through phosphorylation of Nuclear transcription factors
   Many non-steroid hormones use a different mechanism
   For instance Insulin
      **Figure 12-15**
      Binds to cell surface receptor
      Through a series of phosphorylation events, phosphorylated nuclear DNA binding protein
      Alters is interactions as a transcription factor
   Several other hormones use similar mechanism

I. Many Eukaryotic mRNA’s subject to translational repression
   In prokaryotes transcription and translation tightly linked
   In eukaryotes is separate
      So there is a time lag
      And much more opportunity to control steps in between

   If want immediate increase in protein levels, can get faster response if relieve a suppression on an mRNA that is already in cytoplasm

   Seems to be important in several very long genes

   In others seems to be a way of fine tuning

   Also can be used in development

   Only way of control in anuclear cells

Four major mechanisms
   1. Phosphorylation of initiation factors acts as a general suppressant of cellular translation
2. Some proteins bind to 3' end in non-translated region (3'UTR)
   Either bind to translation initiation factors or to 40S ribosome
   to suppress translation
   See for instance 28-34 compared to 27-28

3. Bind proteins that binds with eIF4E and interferes with
   association with eIF4G
   (Eukaryotic initiation factors)
   Again a general suppression of translation

4. RNA mediated regulation
   Will examine in detail next

J. Post-transcriptional Gene Silencing
   Happens in higher Eukaryotes
   Plants and animals higher than nematodes
   Small Additional pieces of RNA called micro-RNA's (miRNA)
   Interact with mRNA
   Often by binding in 3' UTR (untranslated region)
   ie region of RNA between stop codon and
   Physical end of mRNA (poly A tail)
   Bind to make double stranded RNA
   Can speed degradation of mRNA
   Can block translation of mRNA
   In either case mRNA is not translated into protein
   Called Gene Silencing, since is no longer expressed
   1000's of sequences have been identified
   May affect regulation of 1/3 of mammalian genes

   Used in plants as defense against RNA viruses
   (Necessary because no immune system)

   Because many of the miRNA's are present only briefly during
   development, they are sometimes called Small temporal RNA's
   (stRNA's)

Figure 28-35
   Usually synthesized as pieces about 70 bases
   Have lots of hairpin and self complementarity
   Cleaved by endonucleases (one family called ‘dicer’ anther ‘Drosha’)
   Becomes short duplexes about 20-25 base pair long
   These are called small interfering RNA's (siRNA's)
   Lose ½ of duplex
Other ½ binds to mRNA to silence it. 
Then it is not translated or is destroyed
Some miRNA’s interact with only one gene, some with multiple mRNAs so part of a regulon

It may be possible to use this technique medically
If you have a gene you want to silence
Make short pieces of duplex RNA where one strand is complementary to mRNA you want to silence

Add dicer to cleave down to siRNA’s
Inject into cell and let it silence the gene
This method called **RNA interference (RNAi)**
Used in plants as a defense mechanism
Can use on Nematodes (worms)
Just feed them functional RNA’s
They digest it, and partially degrade it
And it silences that gene in the worm!

Method has been used in lab to block HIV and polio infections
So watch this method in the next few years!

**K. Other forms of RNA-mediated regulation in Eukaryotes**
Have now seen several different RNA with functions other than m,r, and t
Call these RNA’s **ncRNA, for non-coding RNA**
Mammalian genome may actually have more ncRNA than coding RNA
So still discovering new uses and methods of control
Some RNA’s bind to proteins to affect their function

Heat shock response in human cells
Heat shock protein 1 (HSF-1)
In nonstressed cell
Monomer
Bound by chaperone Hsp90
Under stress
Released from Hsp90
Forms trimer
Trimer binds to DNA
Activates proteins to respond to stress
A ncRNA of about 600 nucleotides
Stimulates trimerization and DNA binding

Other ncRNA’s known to bind to PolIII to affect activity
L. Development if controlled by a cascade of regulatory proteins

The development of a zygote into a multicellular organism is a real trick changed in cell morphology and protein expression are tightly controlled.

More genes expressed in early development than in rest of cells life
- Sea urchin oocyte - 18,500 different mRNA’s
- In a differentiated cell estimate only 6,000 mRNA’s

Several model systems
- Nematodes, fruit flies, Zebra fish, mice and the plant arabidopsis

Studies of fruit fly (Drosophila melanogaster) are well along so what will be discussed here

Fruit Fly Development
- Life cycle figure 28-36
- Contains several larval stages separated by molts
- Contains metamorphosis from pupa to adult

Important characteristics of embryo
- *Polarity* - distinguish front from back end
- *Metamerism* - separation of body into distinct segments
  - Segment become body parts like head, thorax, abdomen
  - Each segment will have distinct appendages

Have gone a long way to figuring out gene regulating these body patterns

Figure 28-37
- Egg and 16 nurse cells surrounded by layer of follicle cells
- As egg cell formed
- Before fertilization
  - mRNA and proteins from nurse cell and follicle cells deposited in egg cell. Some are going to be important

After Egg is fertilized and laid
- Nucleus divides, and continue to divide in synchrony every 6-10 minutes
- No nuclear membranes distributed in egg cytoplasm

- Between 8th and 11th division
  - Nuclei move to periphery of cell
- After a few additional divisions
PM invaginates to surround nuclei and make layer of cell called bastodern
Now division loses synchrony
Fate of each cell is decided by proteins and mRNA left by nurse and follicle cells

Terminology to be used here
Morphogen - a protein that causes a cell to take up a particular shape or morphology
Morphogens are products of pattern regulating genes

Three major classes of pattern regulation genes that function at different stages in development
Maternal genes
Expressed in unfertilized egg - remain dormant till fertilization
Provide most proteins needed in very early development
Some provide early spatial organization of polarity
Segmentation genes
Transcribed after fertilization
Direct formation of proper # of body segments
3 sub classes
Gap genes - divide embryo into several broad regions
Pair-rule genes -
Segment polarity genes
Pair rule and polarity genes together define 14 stripes that will become 14 segments
Homeotic genes
Expressed later - define appendages will develop in segment

Many regulatory genes in each class
Embryogenesis take about a day
These proteins expressed only in first 4 hours
Regulation at both transcription and translation is occurring
Maternal genes

Some are expressed in nurse cells, some in follicle cells some in egg itself
These genes establish two axes
    anterior-posterior (front back)
    Dorsal-ventral (up down)
Key event is to establish mRNA and protein gradients along axes
Some maternal mRNA’s make proteins and protein diffuse
    Creates asymmetric distributions
Different cells in blastoderm inherit different amounts of protein
    Sets cells on different developmental paths
Products of maternal mRNA include
    Transcription activators and repressor
    Translational repressors
    All used to regulate expression of gene that themselves act
    as pattern regulators

Anterior-posterior axis defined (at least in part) by nanos and bicoid genes

    Bicoid makes anterior (front)
    Nanos - posterior (back)

Figure 28-39

Bicoid

    synthesized by nurse cells and deposited in egg
    Translated soon after fertilization
    Makes concentration gradient with high end marking
    front end

    is a transcription factor that activates transcription of a
    number of other protein involved in segmentation

    Is also repressor for other genes

    Has effect only when Bicoid is above some threshold
    level

    If mess with bicoid levels get very funny
devlopments

Nanos similar

    mRNA deposited at posterior end
    Peak level of protein define tail end
    Is also a repressor
Other the mRNA for other genes like Pumilio, Hunchback and Caudal are uniformly distributed, translation linked to Nanos or Bicoid (see figure)

Segmentation Genes
- Gap genes
- Pair-rule genes
- Segment polarity genes

Three subclasses
- Activated at successive stages of embryo development
- Some gap genes influenced by maternal genes

Homeotic Genes
- Loss of homeotic gene causes normal appendage to appear in the wrong place

- Often very large genes
  - Ubx for instance 77,000 bp up to 50,000 is intron
  - Can take an hour just to transcribe
  - Time to transcribed may be part of control of expression

- Exact way these genes relate to human development is unknown
  - But some regulatory protein are highly conserved 1AA difference between fruit fly and mouse

**M. Stem Cells**
- Adult human many different tissues
  - In each tissue cells are ‘terminally differentiated’
    - I.e. no longer divide
  - If damaged or lost cannot be replaced
- If can regenerate by controlling development could help millions

Key is **stem cells**
- Cells that still have capacity to regenerate

**Figure 28-43**
- After egg is fertilized
  - First few cell division make a ball of cells
  - This ball is totipotent
  - Can differentiate into any tissue or complete organism
  - Continued division lead to blastocyt
    - Hollow ball of cells
      - Outer layer of cells will become placenta
      - Inner layers will become germ layers of developing fetus - ectoderm, mesoderm, endoderm
At this stage cells are Pluripotent
   They can give rise to many different tissues, but not a complete organism
It is these embryonic stem cells that are currently used in research

There are also adult stem cells
   Like hematopoietic stem cells of bone marrow
   But these cells are multipotent
      Can differentiate into many type of blood or bone cells
      But cannot turn into cell of any other tissue
   Only fit into one niche tissue

So there is the problem
   Adult stem cells - hard to isolate, limited use in regenerating tissue
   Embryo stem cells - great potential for differentiation - but ethics of destroying an embryo to get them is real issue

Identification and culturing of pluripotent stem cells from both human and mouse blastocysts has been done since 1998, so we have the start, but still a long way before you can differentiate into a useful tissue
In 2007 actually established limited success in reversing differentiation and taking skin cell to pluripotent stage so some hope there

Do you care to take up the challenge?