Biol 205 Regulation of Transcription
The lac operon: a paradigm of beauty and efficiency
Facing up to eukaryotic cells

Reading Assignments:
Chapter 8: Control of Gene Expression
Pgs. 267-280; Figure 8-15 Look at questions 8-1, 8 & 10

OPTIONAL BUT RECOMMENDED Chapter 5: DNA and chromosomes
Browse through Pgs. 183-191. Look carefully at Figures 5-24

Prokaryotes:

genome size: ?
gene number: ?

Eukaryotes

single celled ??
complex eukaryotes: ??

What fraction of protein-coding genes are being transcribed and translated at any one time or in a given cell type?

Table 1–1 Some Genomes That Have Been Completely Sequenced

SPECIES	SPECIAL FEATURES	HABITAT	GENOME SIZE (1000s OF NUCLEOTIDE PAIRS PER HAPLOID GENOME)	ESTIMATED NUMBER OF GENES CODING FOR PROTEINS
BACTERIA				
Mycoplasma genitalium	has one of the smallest of all known cell genomes	human genital tract	580	468
Synechocystis sp.	photosynthetic, oxygen-generating (cyanobacterium)	lakes and streams	3573	3168
Escherichia coli	laboratory favorite	human gut	4639	4289
Helicobacter pylori	causes stomach ulcers and predisposes to stomach cancer	human stomach	1667	1590
Bacillus anthracis	causes anthrax	soil	5227	5634
Aquifex aeolicus	lithotrophic; lives at high temperatures	hydrothermal vents	1551	1544
Streptomyces coelicolor	source of antibiotics; giant genome	soil	8667	7825
Treponema pallidum	spirochete; causes syphilis	human tissues	1138	1041
Rickettsia prowazekii	bacterium most closely related to mitochondria; causes typhus	lice and humans (intracellular parasite)	1111	834
Thermotoga maritima	organotrophic; lives at very high temperatures	hydrothermal vents	1860	1877
ARCHAEA				
Methanococcus jannaschii	lithotrophic, anaerobic, methane-producing	hydrothermal vents	1664	1750
Archaeoglobus fulgidus	lithotrophic or organotrophic, anaerobic, sulfate-reducing	hydrothermal vents	2178	2493
Nanoarchaeum equitans	smallest known archaean; anaerobic; parasitic on another, larger archaean	hydrothermal and volcanic hot vents	491	552
EUCARYOTES				
Saccharomyces cerevisiae (budding yeast)	minimal model eucaryote	grape skins, beer	12,069	~6300
Arabidopsis thaliana (Thale cress)	model organism for flowering plants	soil and air	~142,000	~26,000
Caenorhabditis elegans (nematode worm)	simple animal with perfectly predictable development	soil	~97,000	~20,000
Drosophila melanogaster (fruit fly)	key to the genetics of animal development	rotting fruit	~137,000	~14,000
Homo sapiens (human)	most intensively studied mammal	houses	~3,200,000	~24,000

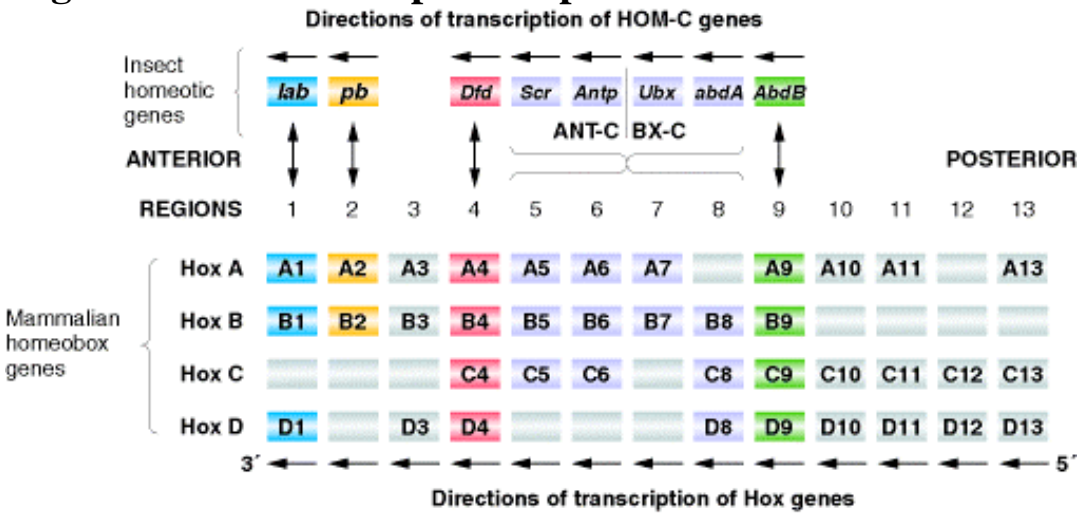
Genome size and gene number vary between strains of a single species, especially for bacteria and archaea. The table shows data for particular strains that have been sequenced. For eucaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Table 1-1 Molecular Biology of the Cell 5/e (© Garland Science 2008)

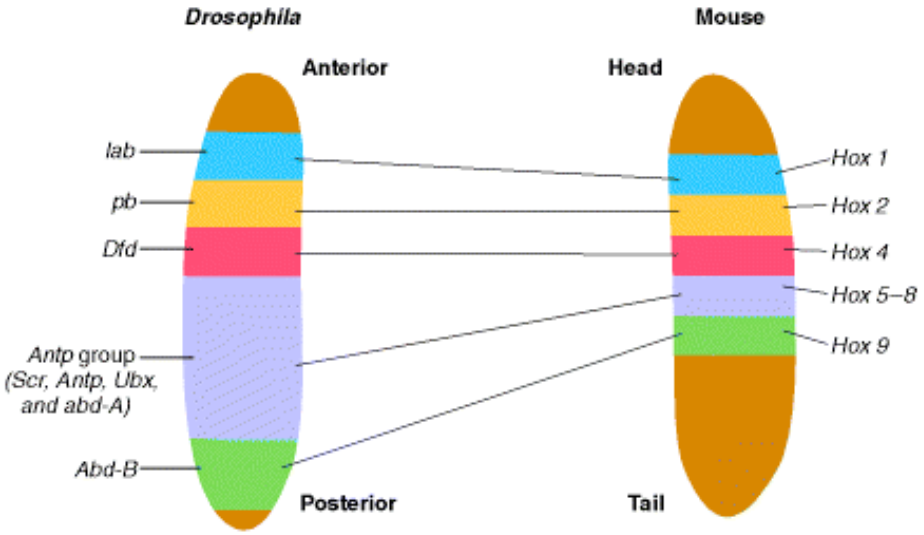
Some interesting biological questions related to differential gene expression in eukaryotes

1) How do you know your head from your tail (the HOX gene complex)

Control of spatial (region and tissue specific expression) expression of genes as well as temporal expression



(a)



(b)

2) Differential gene expression is essential to understanding WHY cloning mammals (using donor nuclei from adults) doesn't work very well

Meet (left to right):
Rainbow, Allie and cc (carbon copy) who is a *genetic clone* of Rainbow. Allie is cc's surrogate mom

Enough experience has accumulated to assess the risks of vertebrate cloning:

- animal cloning is inefficient and is likely to remain so for the foreseeable future
- It took 276 unsuccessful attempts before the sheep Dolly was produced
- The cloned cat CC was the product of 188 cloning attempts (188 enucleated eggs plus cumulus cell nucleus) ➔ 82 embryos ➔ one live kitty
- cloning results in gestational and neonatal developmental failures
- at BEST, a few percent of the nuclear transfer embryos survive to birth and, of those, many die within the perinatal period
- newborn clones often display respiratory distress and circulatory problems and even apparently healthy survivors may suffer from immune dysfunction, or kidney or brain malfunction

Many of the cloning failutes and many of the fetal abnormalities and abnormalities in those few clones that are born alive probably result from failures in ***genomic reprogramming which causes abnormalities in gene expression****.*

3) How come we're not chimps?

It is well established that at least 98.7% of human DNA (coding and non-coding is identical to that of chimpanzees).

Which of our genes makes us human?

IN other words what genes have we acquired in the last ********* ***years that the common ancestor of chimps and human did not have?***

Maybe is this not the correct question?

Differential gene expression of our almost identical genomes may be the reason that you are a human and not a chimp

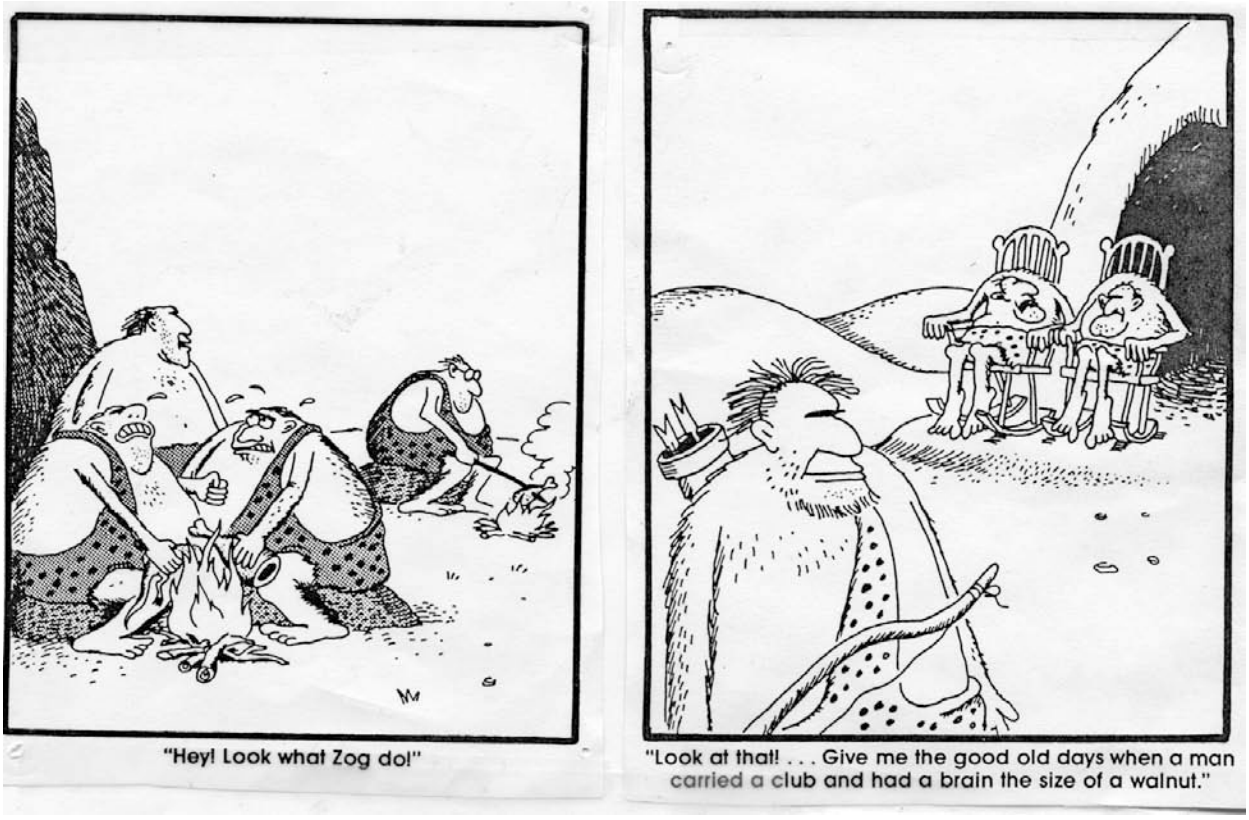
PRIMATE EVOLUTION:

Gene Activity Clocks Brain's Fast Evolution

Elizabeth Pennisi

A team of molecular biologists has taken a stab at defining what makes us human. Its answer: We're set apart from other primates not so much by differences in the makeup of our genes but by relatively recent changes in how active those genes are. Such changes are most dramatic in the brain, where they've occurred at a faster rate in humans than in other primates, they report on page 340.

Science 296: 233 April 12, 2002



Intra- and Interspecific Variation in Primate Gene Expression Patterns

Wolfgang Enard,[1*] Philipp Khaitovich,[1*] Joachim Klose,[2] Sebastian Zöllner,[1] Florian Heissig,[1] Patrick Giavalisco,[3] Kay Nieselt-Struwe,[4] Elaine Muchmore,[5,6] Ajit Varki,[5] Rivka Ravid,[7] Gaby M. Doxiadis,[8] Ronald E. Bontrop,[8] Svante Pääbo[1†]

Although humans and their closest evolutionary relatives, the chimpanzees, are 98.7% identical in their genomic DNA sequences, they differ in many morphological, behavioral, and cognitive aspects. The underlying genetic basis of many of these differences may be altered gene expression. We have compared the transcriptome in blood leukocytes, liver, and brain of humans, chimpanzees, orangutans, and macaques using microarrays, as well as protein expression patterns of humans and chimpanzees using two-dimensional gel electrophoresis. We also studied three mouse species that are approximately as related to

Science 296: 340 April 12, 2002

Facing up to eukaryotic organisms:

differentiation of a complex array of cells and tissues from a single cell

What does molecular biology tell us about this process?

The final outcome: variety of tissues producing a variety of specific gene products located in the proper orientation to each other

How has the study of prokaryotic model organisms helped us to understand the more complex gene expression processes in eukaryotes?

The E. coli genome projects and follow-up studies on gene expression have provided some real data:

How many protein-coding genes does E. coli have?

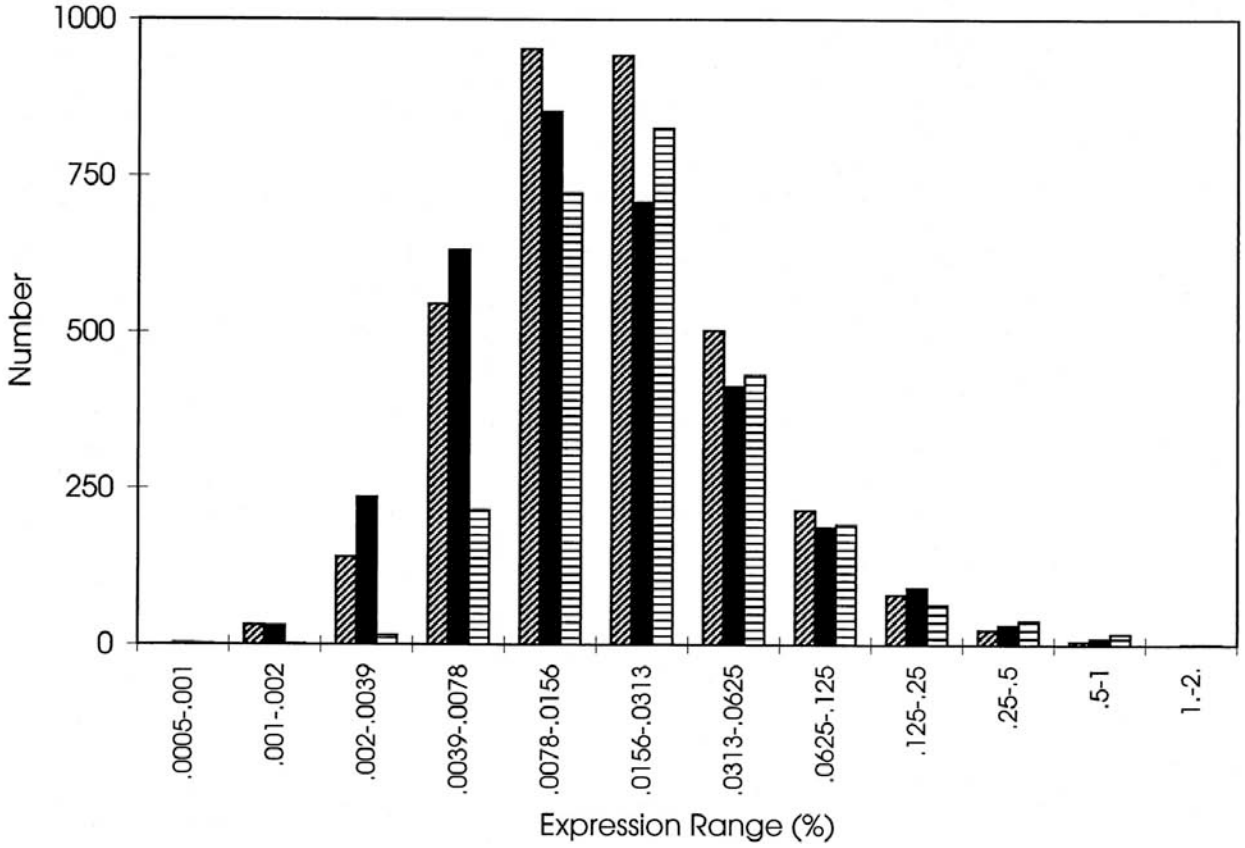
➔ 4,290

What fraction of the genome is being transcribed at any one time?

From a paper entitled: *High-density microarray-mediated gene expression profiling of E. coli*

What fraction of the genome is transcribed?

j. bacteriology Jan 2001 pg. 545-556



Distribution of expressed genes. The histogram plots number of genes as a function of expression range. Distributions observed in RNAs derived from cell:

Diagonally striped bars: growing exponentially in minimal medium

Solid bars: cells transitioning to the stationary phase in minimal medium

Horizontally striped bars: cells growing exponentially in rich medium

Expression of 766, 1,030, and 1,776 genes was not detected under the three respective conditions.

Transcription data:

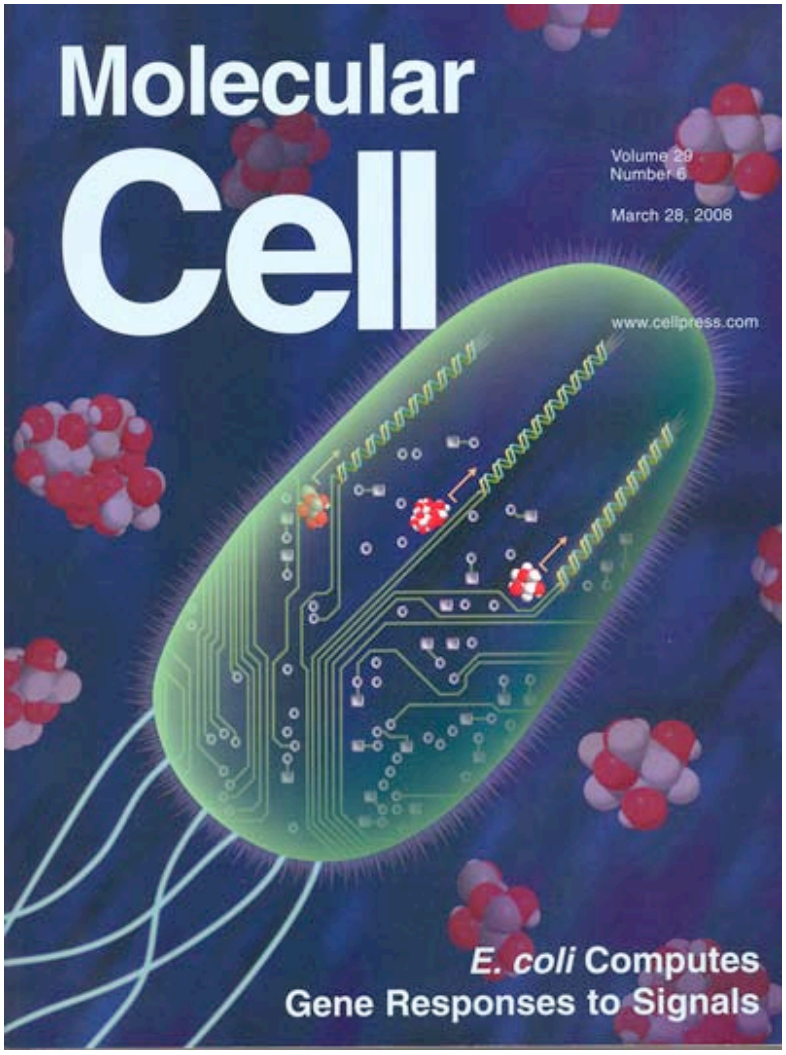
minimal media (carbon/energy source such as glucose, inorganic salts and water): 82% of genes have detectable levels of transcription

rich media (carbon/energy sources, amino acids, vitamins, nucleotides, fats, etc): 59% of genes have detectable levels of transcription

➔ If you've only got a few thousand genes, why not synthesize all proteins all of the time at a moderate level?

1. energetically expensive to synthesize proteins
2. levels of some products need to be carefully controlled due to specific requirements of physiology or biochemistry or issues of toxicity
3. some proteins products may be incompatible if in the cell at the same time -- one might prevent the functioning of another
4. all organisms (especially microbes and plants) need to be able to respond to changing environmental conditions by altering their gene expression and under certain environmental conditions a specific protein might be disadvantageous
5. differential gene expression underlies cell specialization and the development of complex multicellular organisms

How has the study of prokaryotic gene regulation helped us to understand eukaryotic gene regulation?



As illustrated by the data shown above, E. coli can grow under diverse environmental conditions by adjusting its gene expression profiles. To efficiently perform this task, the bacteria need to sense and integrate multiple signals and compute by a transcriptional network the desired response for each gene.

Prokaryotic cells:

In the context of prokaryotic cells, control of gene expression is primarily focussed on
(1) responding efficiently to a changing environmental context
(2) controlling metabolic and physiological processes

Prokaryotic Cell specialization/differentiation

Bacillus forms spore under conditions of nutrient depletion
Cyanobacteria produce two types of cells will incompatible biochemical processes:

- photosynthetic, non-N_2 fixing cells
- N_2 fixing, non- photosynthetic cells
  (nitrogenase is O_2 sensitive)

CONTROL OF GENE EXPRESSION means controlling the level of active gene product (this is a broad definition of control of gene expression)

➔ this control can be accomplished at different points in the generation of the protein product:
transcription: generation of the mRNA copy

post-transcription: processing of mRNA -- alternative splicing in eukaryotes

translation: generation of the protein from the mRNA

post-translation: activity of the protein is controlled by one of a number of different types of mechanisms

What are various types of post-translational mechanisms for controlling protein activity?

The *lac* operon:
a paradigm of beauty and efficiency

1. The study of this operon defined two key components of all transcriptional regulatory circuitries:

- *trans-acting regulatory proteins (activators and repressors)*
- *cis-acting sites on DNA which bind the trans-acting proteins*

2. Control of expression of the lac operon occurs at the transcriptional and post-translational levels

3. The lac regulatory circuitry involves allosteric effects that reveal the presence of an extracellular signal

4. The *lac Z* gene is used widely in the lab as a reporter gene

the lac operon

Paradigm:

- **an outstandingly clear or typical example**
- **the prototype**
- **the perfect example**

operon

- a unit of bacterial gene expression and regulation, including structural genes and control elements.
- genes contained in an operon are under coordinate expression: all the genes are expressed (or not expressed) in unison.
- The activity of the operon is controlled by regulator gene(s), whose protein products interact with the control elements of the operon.

lac Z beta galactosidase

lac Y lactose permease

lac A transacetylase

Phenomenology of β-galactosidase induction

Expression of the the β-gal enzyme under various physiological conditions

glucose	lactose	level of β-gal
-	-	1X
+	-	1X
+	+	20 X
-	+	1000 X

How does the cell turn transcription of the lac Z gene on and off?

How does a cell "know" when to turn this gene on or off?

These questions were first addressed in th 1960's by Francois Jacob and Jacques Monod

left: monod middle: jacob

The lac operon circuitry was initially proposed by these two individuals based solely on very elegant genetic analysis of mutant strains

The components of the lac operon regulatory circuitry:

1. cis-acting sites

2. trans-acting regulatory proteins

3. a mechanism for regulating the activity of the regulatory proteins -- a way of "knowing" whether lactose and glucose are available

cis-acting site:

- a site on the DNA that only affects the molecule of DNA on which it resides
- a genetic element that must be on the same chromosome in order to influence a gene's activity
- a cis-acting site does not specify a protein (diffusable activity)
- a cis-acting site is effectively inert until it is contacted by its cognate regulatory protein

cis-acting sites in lac operon
lac promoter
lac operators
cap-binding site

operator: cis-acting site at one end of an operon that acts as a binding site for repressor protein

Trans-acting factors: Positive (activator) and Negative (repressor) proteins:

Rate of initiation of transcription in E. coli depends on:

1. the inherent affinity of RNA polymerase holoenzyme for a particular (sequence specific)
2. the presence of regulatory proteins that bind at or near the promoter and act to increase or decrease the rate of transcription inititation

Negative Regulation: binding of the regulatory protein (repressor) to its operator inhibits transcription

Positive regulation: binding of the regulatory protein (activator) stimulates transcription

Gene specificity is conferred by the presence or absence of a specific cis-acting site at the promoter

Only genes that have the *cis-acting site* that the regulatory protein binds to will be affected by its presence

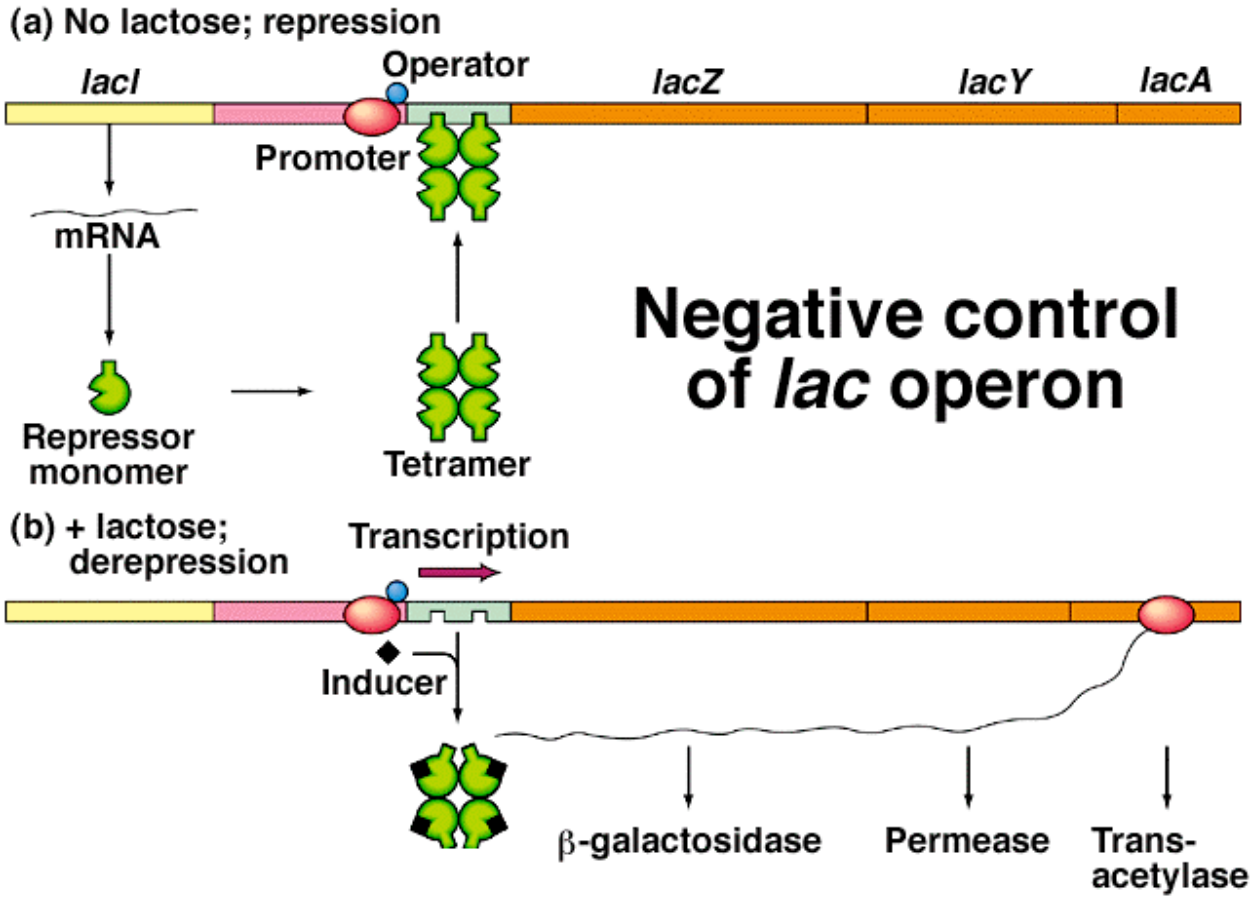
A way to regulate the activity of the regulators

lac I gene:

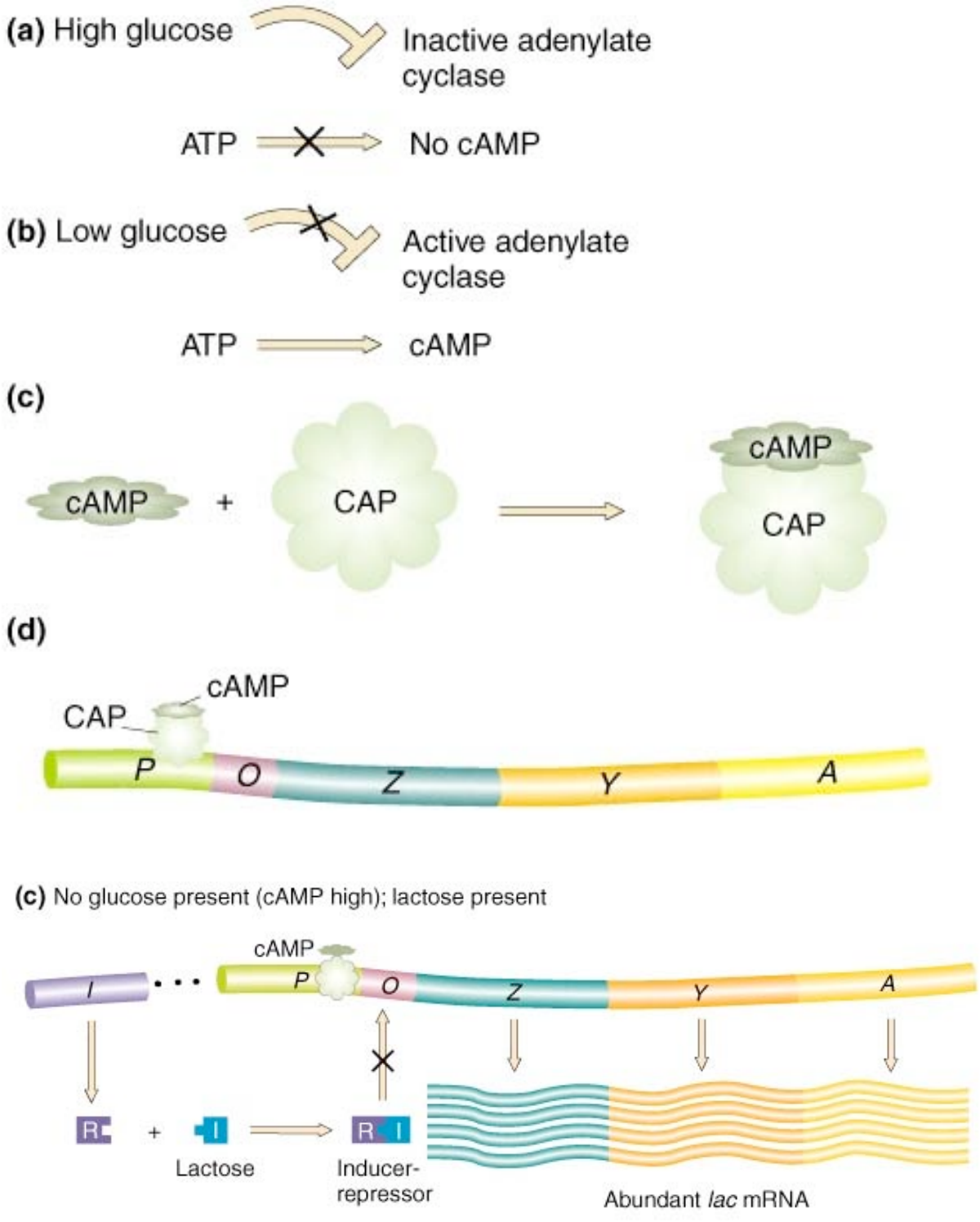
- codes for lac repressor protein (negative regulator) and is not part of the operon per se
- lactose is its allosteric regulator
- cannot bind to operator when bound by lactose at its allosteric site

CAP (cAMP binding) protein:

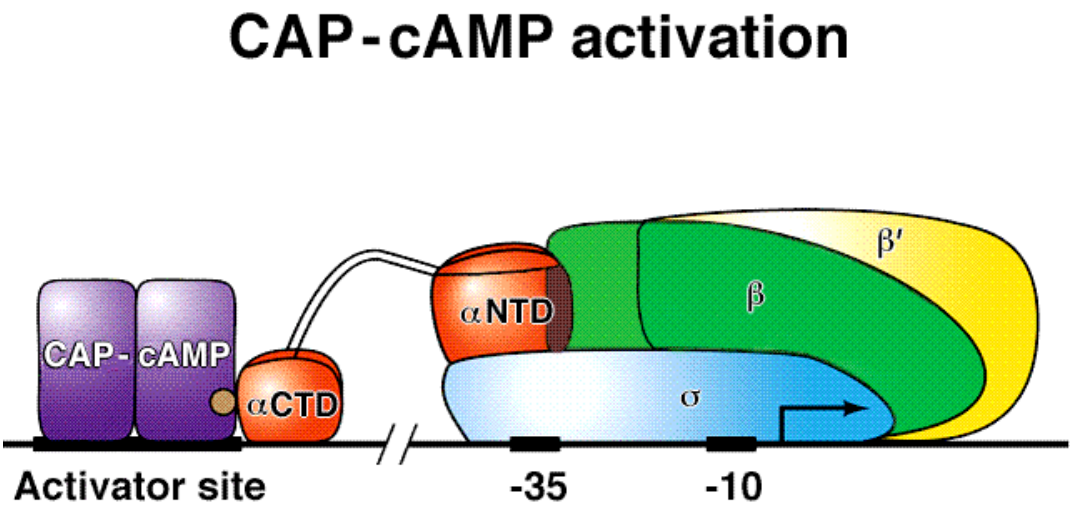
- catabolite activating protein (positive regulator)
- cAMP is its allosteric regulator
- cannot bind to cis-acting site on lac promoter unless cAMP is bound to allosteric site



CAP = cyclic AMP (cAMP) receptor protein



RNA polymerase binds weakly to the lac operon promoter in the absence of the CAP protein



***E. coli* RNA polymerase has four subunits**

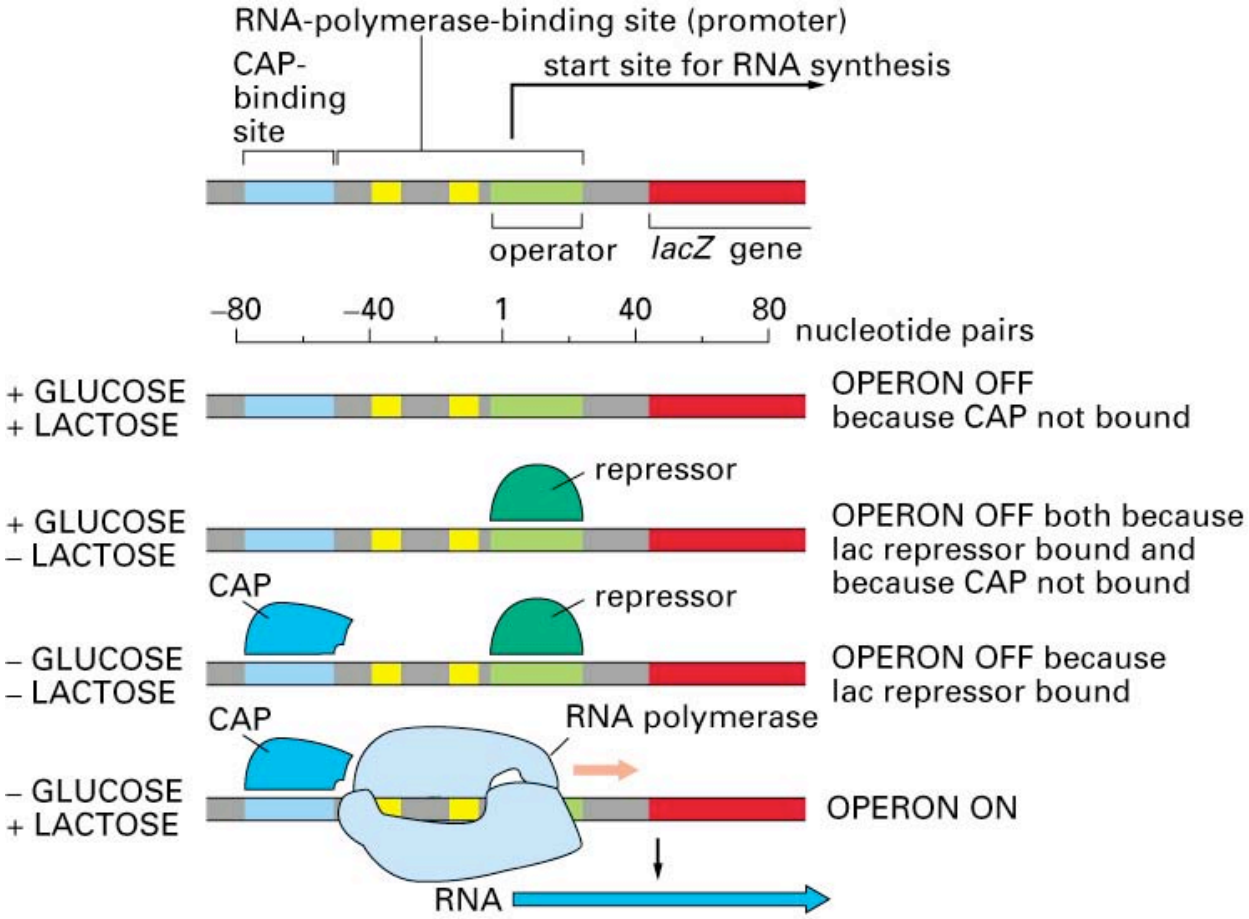


Figure 7–38. Molecular Biology of the Cell, 4th Edition.

Univerisal Themes in Gene Regulation at the level of Transcription

There are two key components of all transcriptional regulatory circuitries:

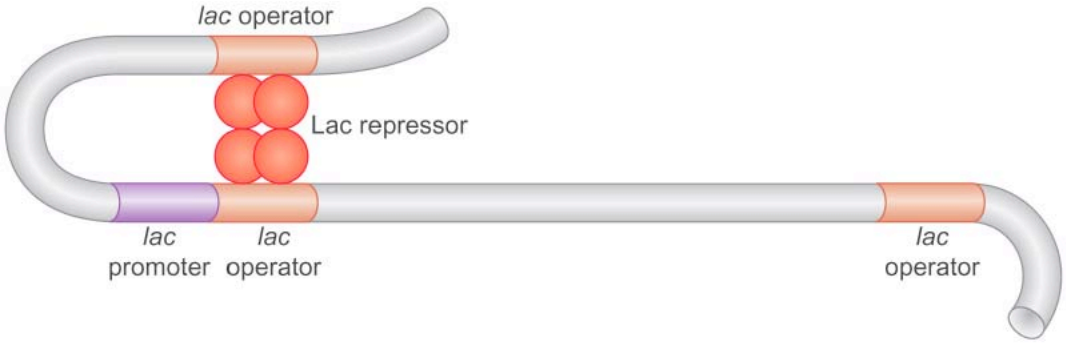
trans-acting regulatory proteins (activators and repressors)
cis-acting sites on DNA which bind the trans-acting proteins

A cis-acting site is INERT until it is contacted by its cognate regulatory protein

The "cognate" proteins contact the corresponding cis-acting site with a ***sequence specific DNA binding domain***

There must be a way to regulate the trans-acting regulatory proteins. [Control of the regulators of the lac operon occurs at what level?]

Binding the the regulatory protein to the DNA involves bending and distortion of the DNA helix



General Themes seen in both eukaryotic and prokaryotic cells

Transcriptional regulatory circuitries in euk cells has the same basic components as prok cells

Regulatory proteins

- trans-acting DNA-binding proteins that recognize and bind to specific cis-acting sites: *repressors, activators, negative regulators, positive regulators, transcription factors*

Cis-acting sites

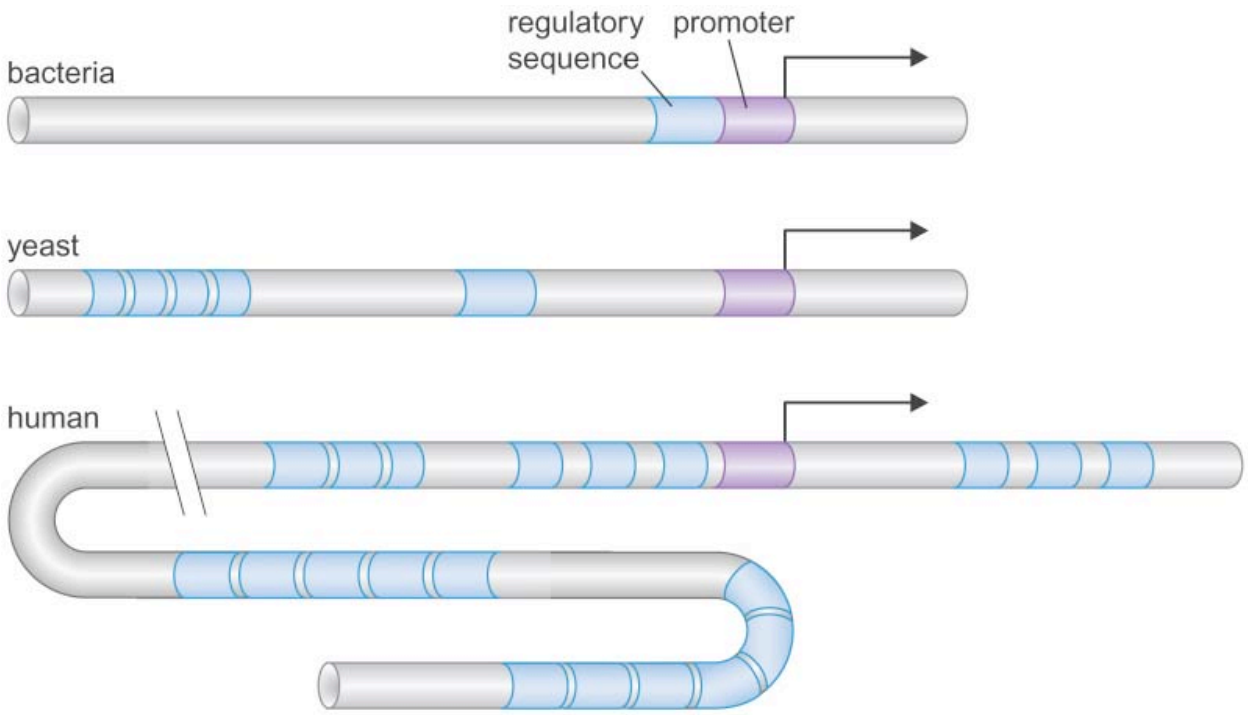
- gene specific -- often (but not necessarily) adjacent to the core promoter region
- known various names such as operators in proks and *enhancers & silencers in euks*
- ***these sites are inert until bound by their cognate regulatory proteins***

A mechanism for regulating the activity of the regulatory protein

- level and/or activity of regulator is modulated in response to some signal
- signal could come from the external environmental, the internal environmental (ie. hormones in multi-cellular organisms), intracellular signals (developmental regulatory proteins)

Also involves

- core promoter sequences (not gene specific)
- generic TFs in eukaryotes (not gene specific) and sigma factor in proks
- RNA polymerase complex



Watson Figure 17-1 The regulatory elements of a bacterial, yeast and human gene

From Nature 421:444 2003 *The digital DNA code*

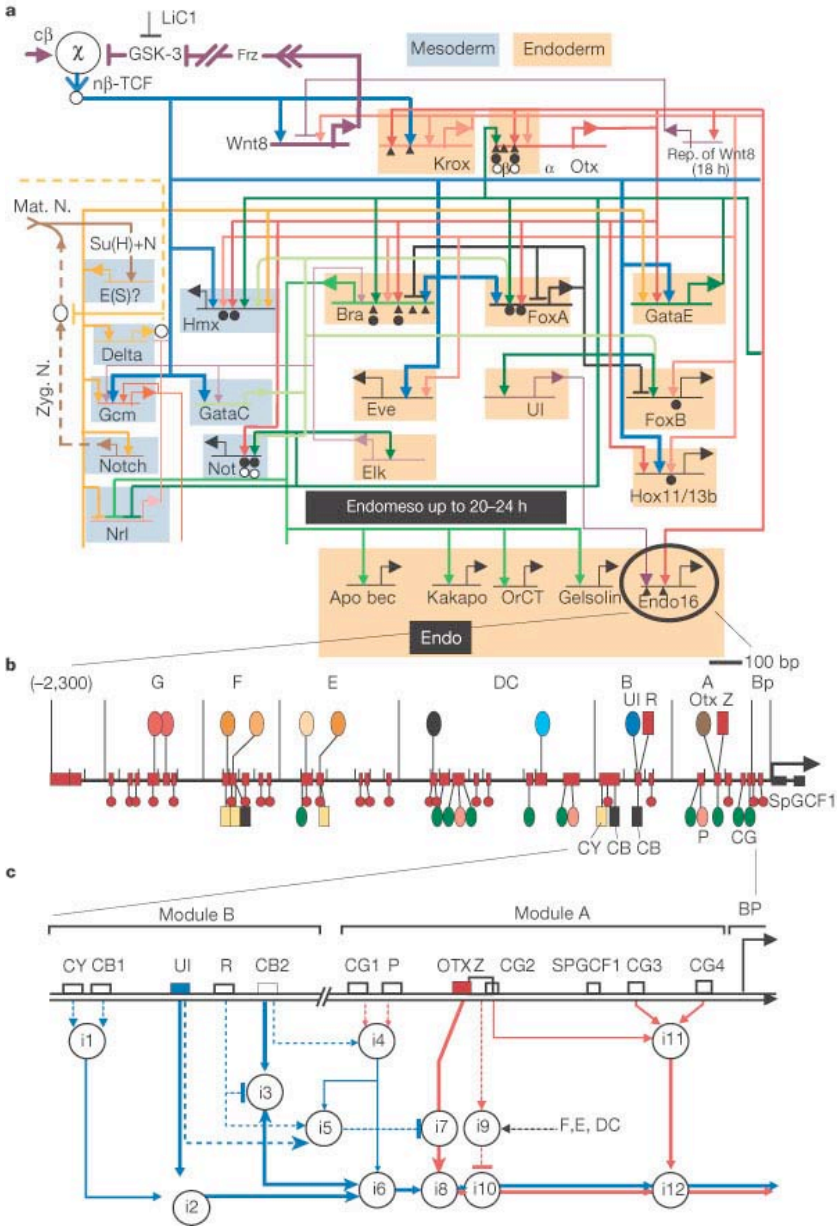


Figure 2 A gene regulatory network involved in sea urchin development16. **a,** Part of the network of transcription factors and their interactions with the control regions of other transcription factors. Genes are indicated by horizontal lines; arrowheads indicate activation; " symbols indicate gene repression. **b,** An enlargement of the promoter region of a gene, called *endo 16*, that helps modulate the development of the endoderm. It contains 34 binding sites (rectangles) for 13 different transcription factors and cofactors (illustrated as rectangles or lollipops, respectively). Six modules (A–G) of transcription factors and binding sites carry out discrete functions to developmentally regulate *endo 16*. **c,** Diagram depicting the logical structures of the A and B control circuits during sea urchin development.

Regulatory Proteins: *Transcription Factors that recognize specific promoter elements or other cis-acting sequences*

These transcription factors have a variety of DNA-binding motifs

- A search of the human genome sequence has revealed more than 2000 genes that encode transcription factors which are involved in the regulation of gene expression
- In this figure they are categorized according their protein "family"
- Assignment to a family is based on specific protein motiffs that relate to how the protein contacts DNA or other aspects of their function

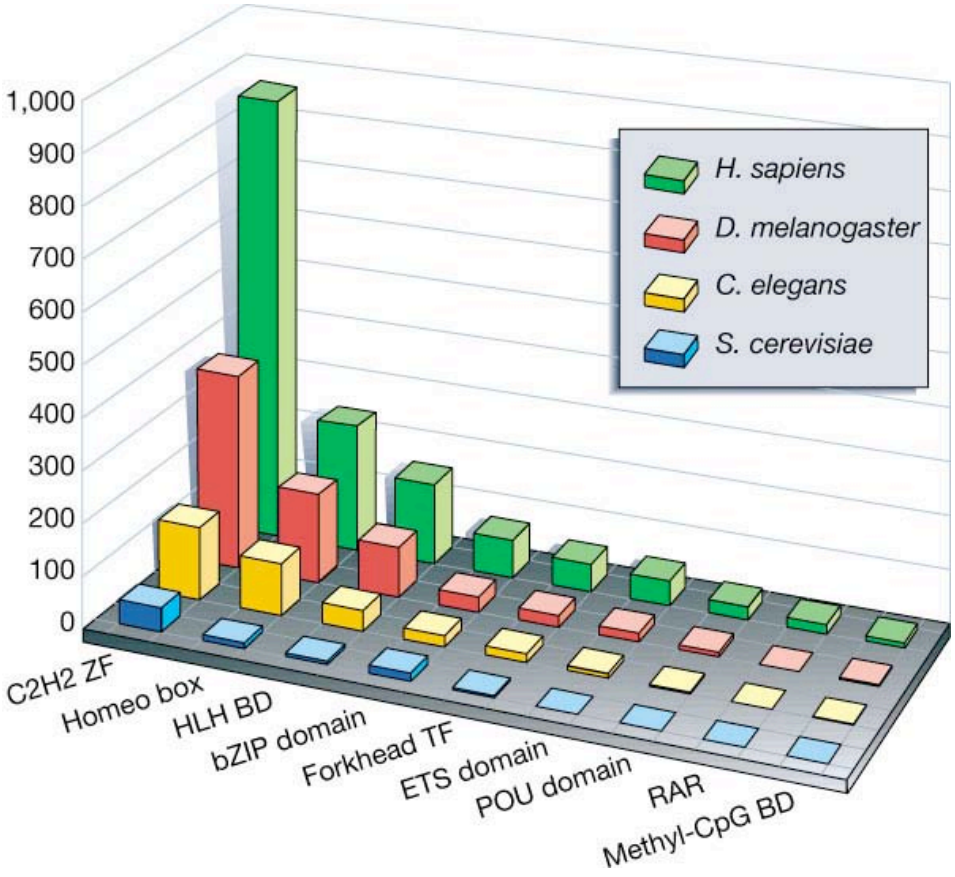
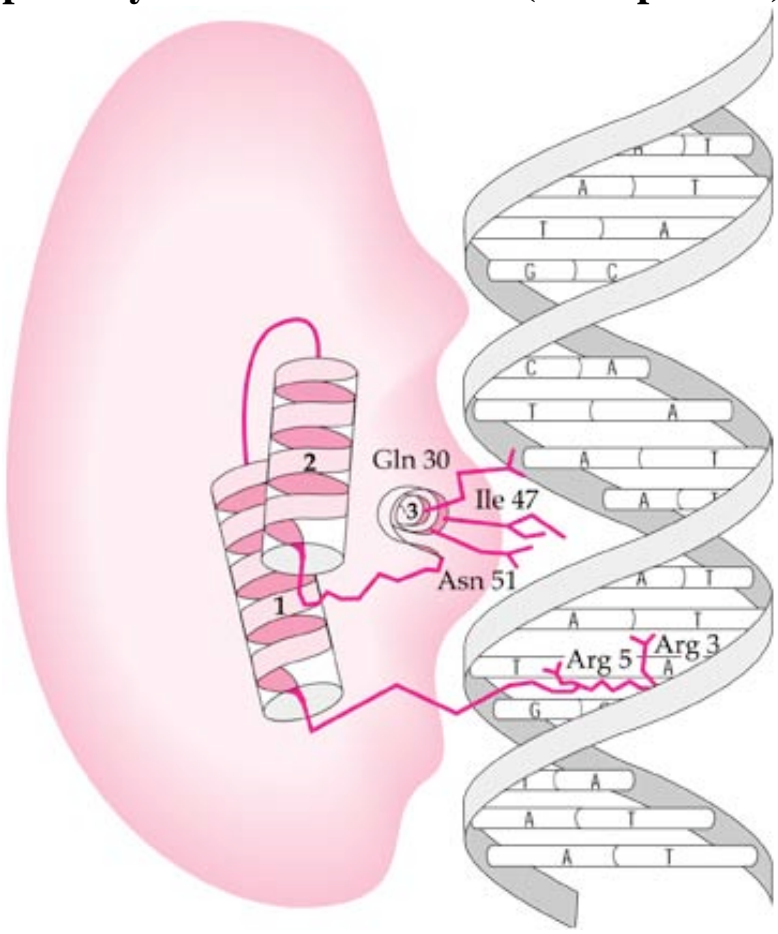


Figure 1 Genome-wide comparison of transcriptional activator families in eukaryotes. The relative sizes of transcriptional activator families among Homo sapiens, D. melanogaster, C. elegans and S. cerevisiae are indicated, derived from an analysis of eukaryotic proteomes using the INTERPRO database, which incorporates Pfam, PRINTS and Prosite. The transcription factors families shown are the largest of their category out of the 1,502 human protein families listed by the IPI. Nature 409:832 Feb. 15, 2001 *Expressing the Human Genome*

Homeodomain (homeobox) DNA binding motif is like the prokaryotic helix-turn-helix (lac repressor)



The homeodomain of the Engrailed protein binds to a particular site in the DNA. Helix 3 contacts the base pairs in the major groove, while the amino-terminal portion of the homeodomain enters the minor groove. (AfterPabo and Sauer, 1992.)

Sequence specific binding of eukaryotic TFs involves H-bonds and other non-covalent interactions

CIS-ACTING SITES IN EUKARYOTIC PROMOTERS

- **CORE PROMOTER** (includes TATA box)

- **PROMOTER-PROXIMAL ELEMENTS:** found withing 100-200 bases of the transcription initiation site

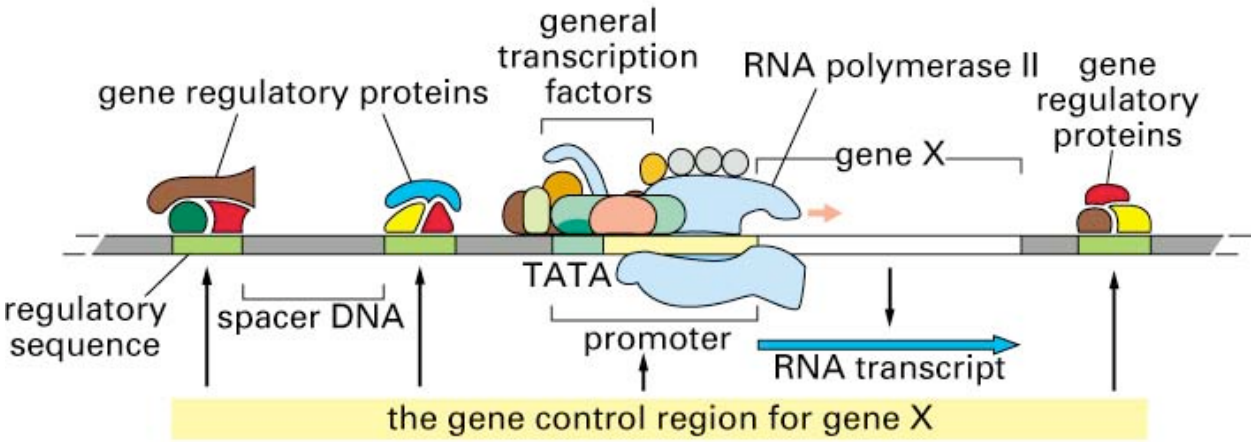
- **DISTANCE-INDEPENDENT SITES:** cis-acting elements that can exert their effects at considerable distance either upstream or downstream from the promoter (includes *enhancers* and *silencers*)

DISTANCE-INDEPENDENT SITES:

enhancers or **silencers** (in higher eucaryotes)
upstream activator sequences (UAS) in yeast

- **gene-specific** (and in multicellular eukaryotes may control tissue specificity of expression -- see below)
- ***can be located 5' of, 3' of, or within an intron of a gene.*** The orientation of these sequences relative to the transcription start site is relatively unimportant.
- Furthermore, such sites may be located at quite a distance (thousands of base pairs) away from the gene that they are regulating.

What features of this promoter complex are not characteristic of prokaryotic cells?



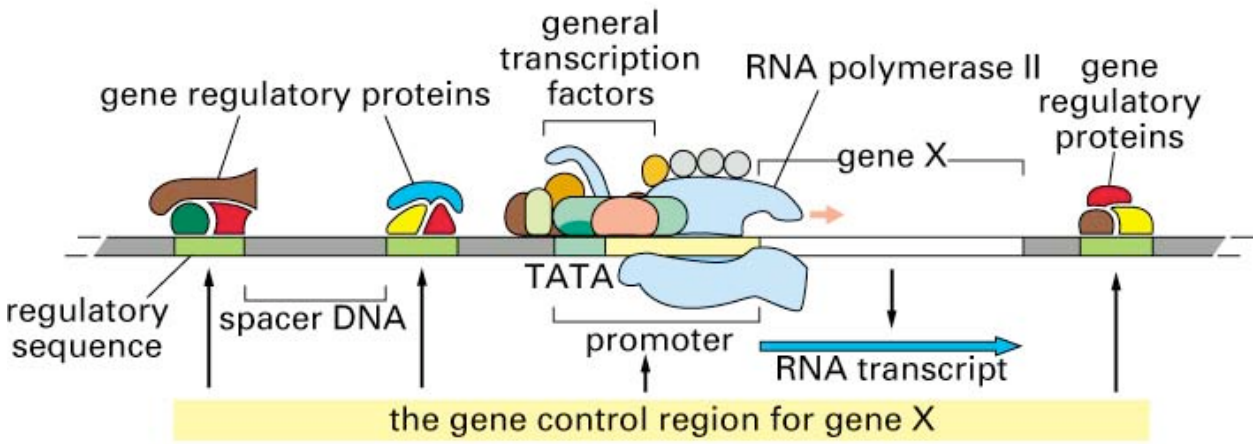
Core RNA polymerases do not recognize promoters on their own

- TFII A, B ,D, E, F & H are basal or general transcription factors used at every eukaryotic promoter
- The action of these factors is required to assist pol II in recognizing promoter sites and initiation transcription -- is there a prokaryotc parallel to the basal transcription factors

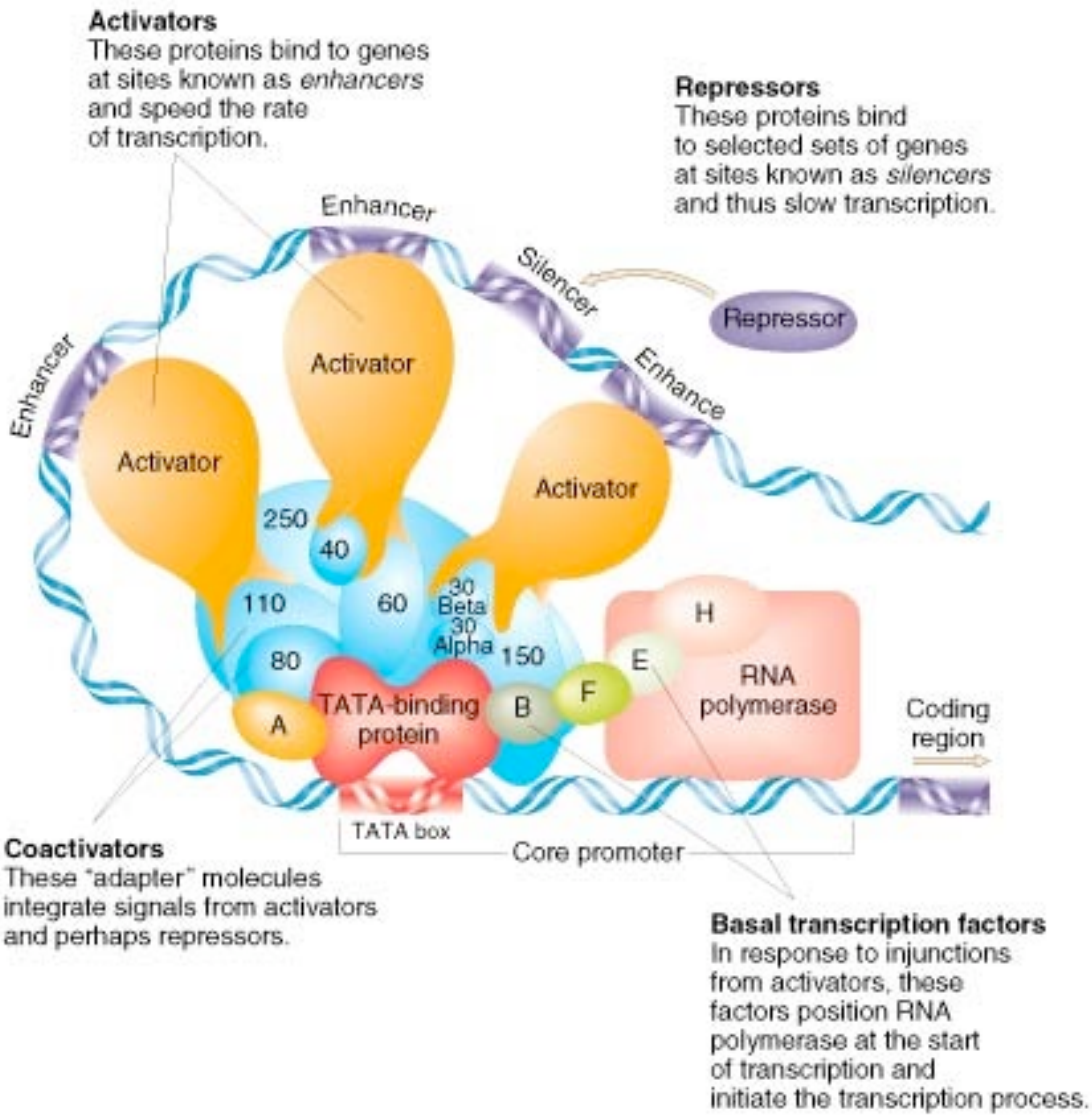
The general transcripton factors are essentially the same for each gene transcribed by pol II

The regulatory transcription factors and their cis-acting sites will vary from gene to gene

What features of this promoter complex are not characteristic of prokayotic cells?



What's wrong with this figure? How can a cis-acting site affect RNA polymerase at a promoter 1000's of bases distant from the site?


Activators
These proteins bind to genes at sites known as enhancers and speed the rate of transcription.
Repressors
These proteins bind to selected sets of genes at sites known as silencers and thus slow transcription.
Enhancer
Silencer
Repressor
Enhancer
Enhance
Activator
Activator
Activator
250
40
110
60
30 Beta
30 Alpha
150
80
A
TATA-binding protein
B
F
E
H
RNA polymerase
Coding region
TATA box
Core promoter
Coactivators
These "adapter" molecules integrate signals from activators and perhaps repressors.
Basal transcription factors
In response to injunctions from activators, these factors position RNA polymerase at the start of transcription and initiate the transcription process.

The activity or availability of a transcription factor can be controlled in many different ways:

- ♦ transcriptional control
- ♦ alternative splicing
- ♦ translational control
- ♦ post-translational control (allostery, phosphorylation, dimer or heterodimer formation, sequestration in a cellular compartment)

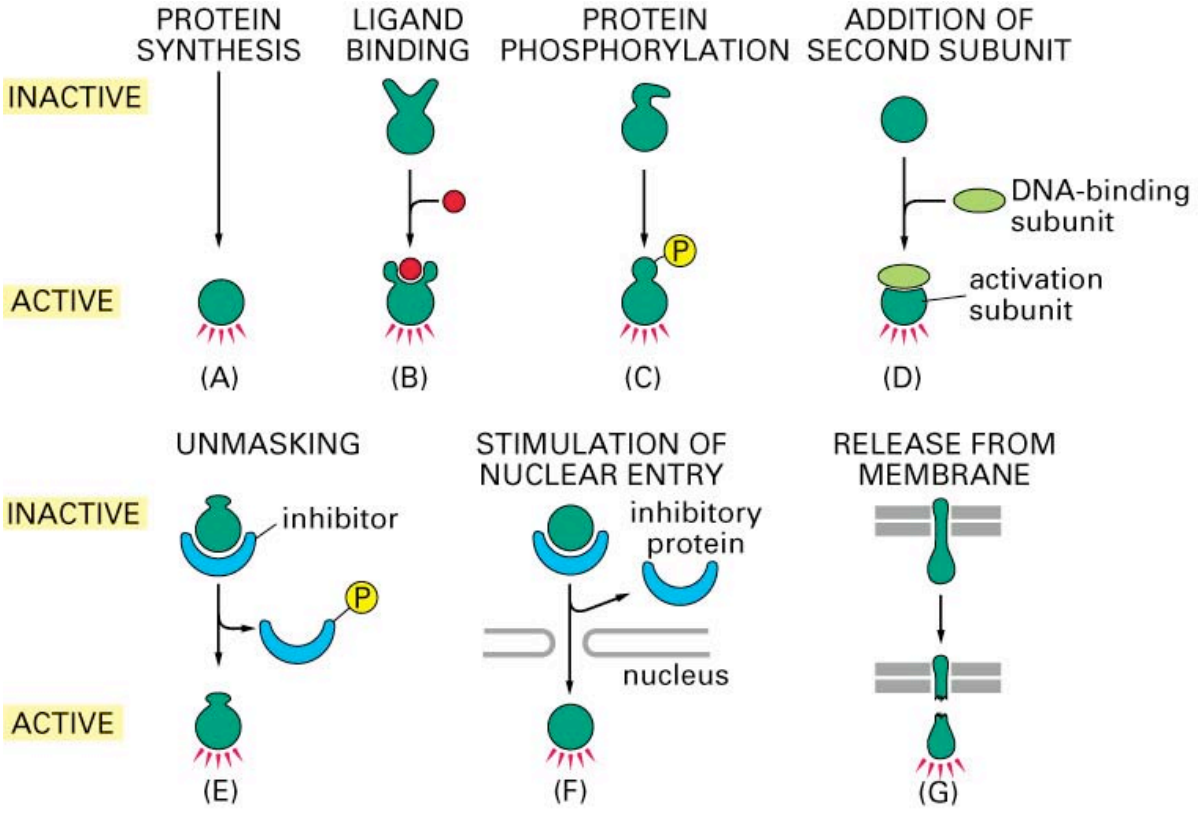


Figure 7–58. Molecular Biology of the Cell, 4th Edition.

Gene Expression in Eukaryotes

Gene ----------> Active protein product

1. Rate of transcription initiation **
2. Alternative splicing patterns
3. Transcript stability
4. Translational regulation
5. Post-translational processing or activation

↓ cis-acting sites ↓ trans acting proteins

RNA pol II complex with promoter:
rate of formation

↑ CpG methylation (complex euks)

↑ chromatin structure (histone modification & nucleosome remodelling)

♦ *Normal development depends on a precise sequence of changes in the configuration of the chromatin and the methylation state of the genomic DNA*

These so-called epigenetic alterations are involved in tissue-specific expression of genes