BI 8 LECTURE 11

QUANTITATIVE ASPECTS OF TRANSCRIPTION FACTOR BINDING AND GENE REGULATORY CIRCUIT DESIGN

Ellen Rothenberg

9 February 2016

Major take-home messages from λ phage system that apply to many other complex systems (including eukaryotes)

- Transcription factors may be bifunctional… depends on context of sites that they bind to

- Binding of multiple copies of a factor to neighboring sites can yield much sharper regulatory responses to transcription factor dose than single binding events… provided that bound factors interact with each other to stabilize (***cooperativity***)

- Biological switches often use mutually antagonistic factors that each enhance their own expression while blocking the others'… causes winner-take-all system behavior

Transcription factors bind nonspecifically to DNA in general, as well as having specific, extra strong recognition for a particular target site

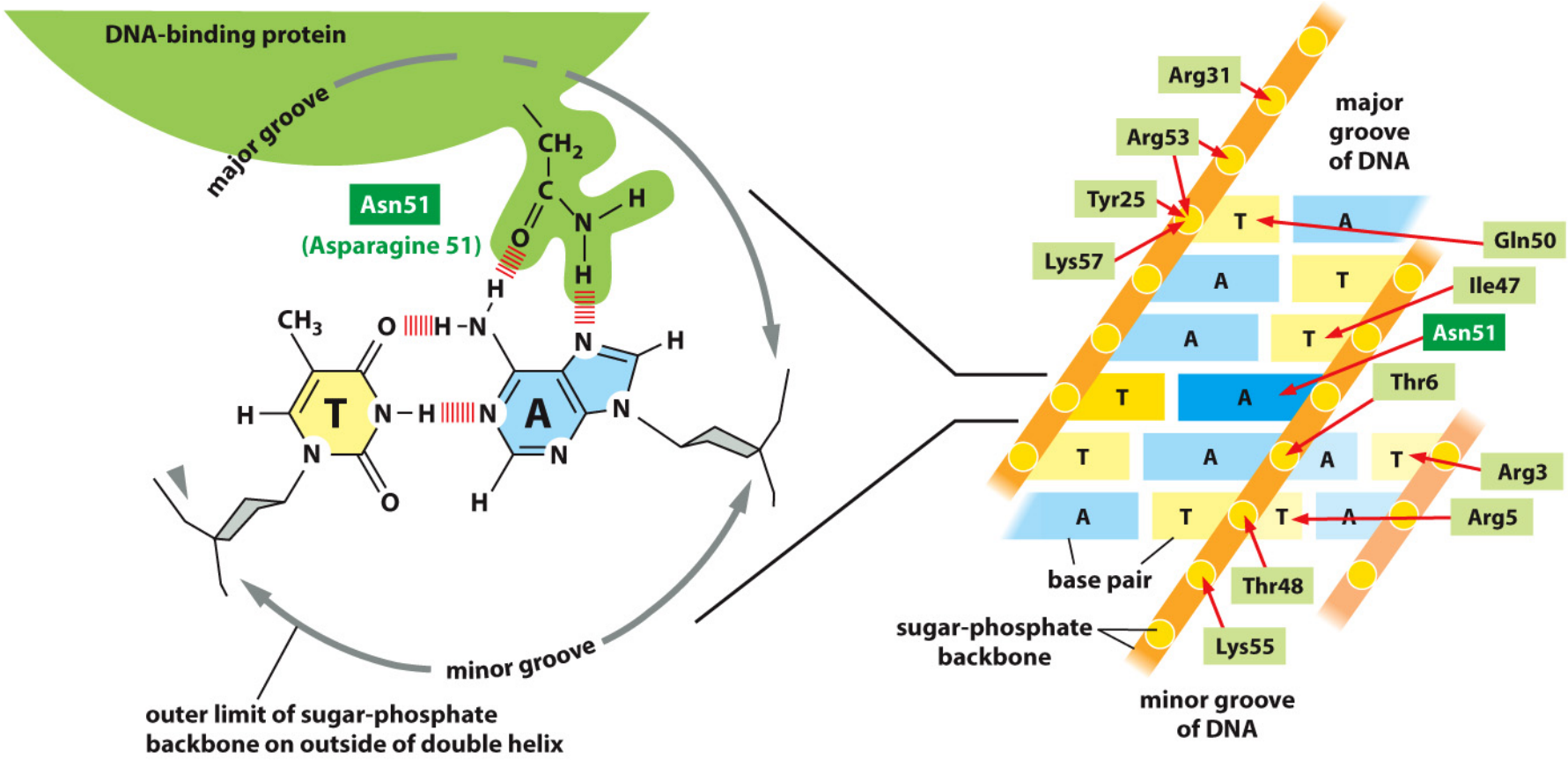


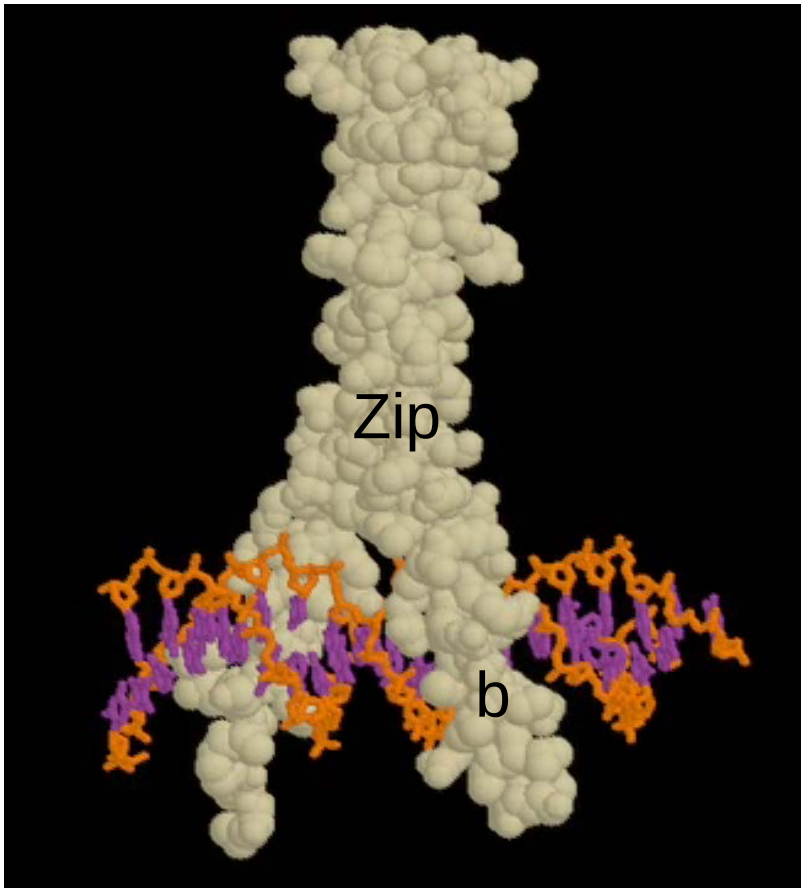
Figure 7-8 Molecular Biology of the Cell 6e (© Garland Science 2015)

Transcription factors with the Leucine zipper motif -- a famous coiled coil with strong dimerization power – use a ***different*** part of the structure to bind to the DNA itself: many backbone contacts

(side)

(top)

Leucine zipper itself



For structures of many different classes of TFs, see Panel 7-1, Alberts et al. (2015)

Thinking about “recognition” as an aspect of binding: Equilibria between alternative states are the “constraints” but also the tools of biology

- For aggregates of individual molecules, equilibria between synthesis and degradation
- For reactions, between forward and reverse reaction
- For chemical recognition, between bound and free

Most binding, in particular, is dynamic, with molecules binding and unbinding at different rates

- For transcription factors, rate of binding to naked DNA is fast, not rate limiting
- High affinity binding is different from low affinity binding because of a slower UNBINDING rate

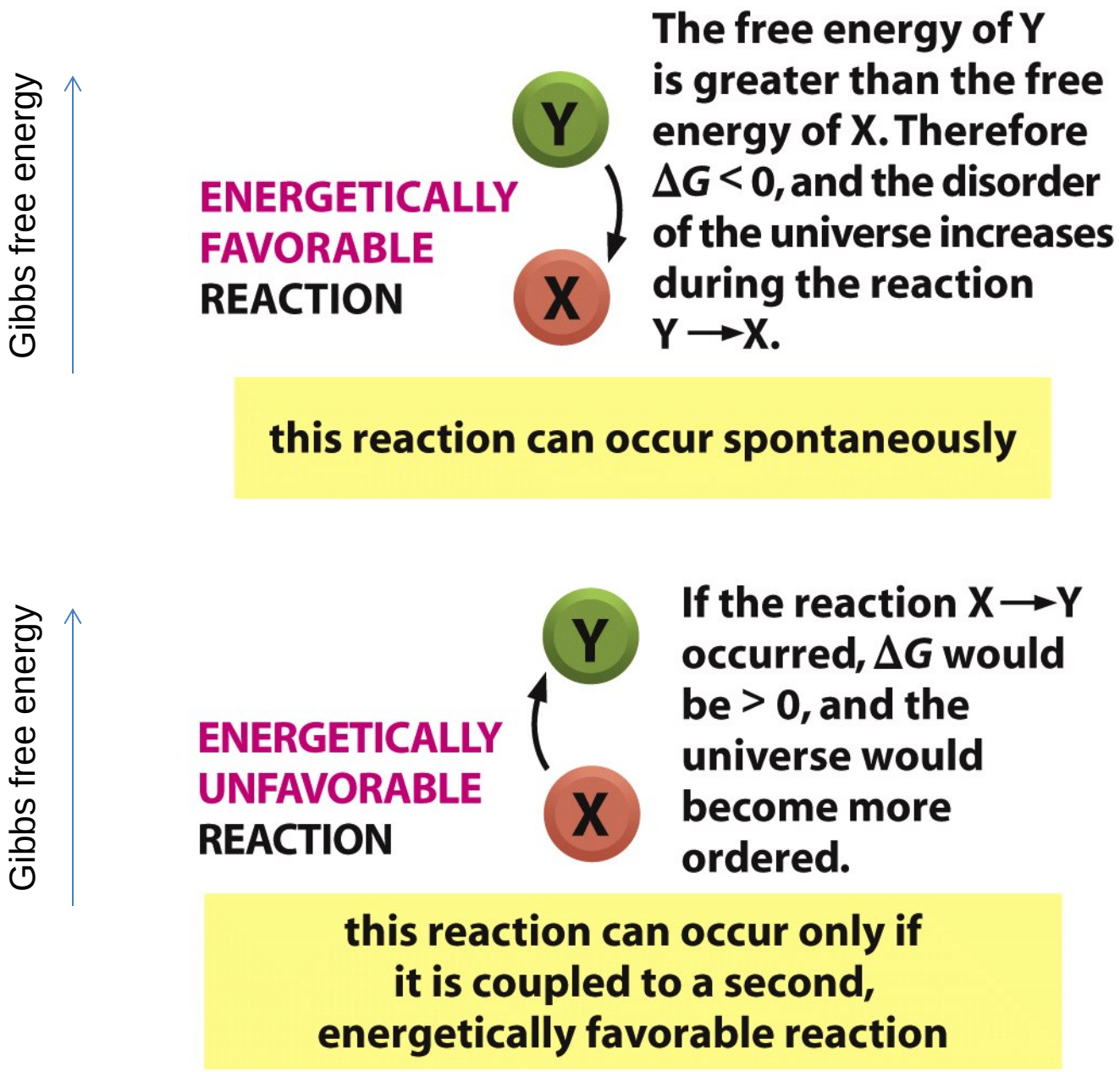
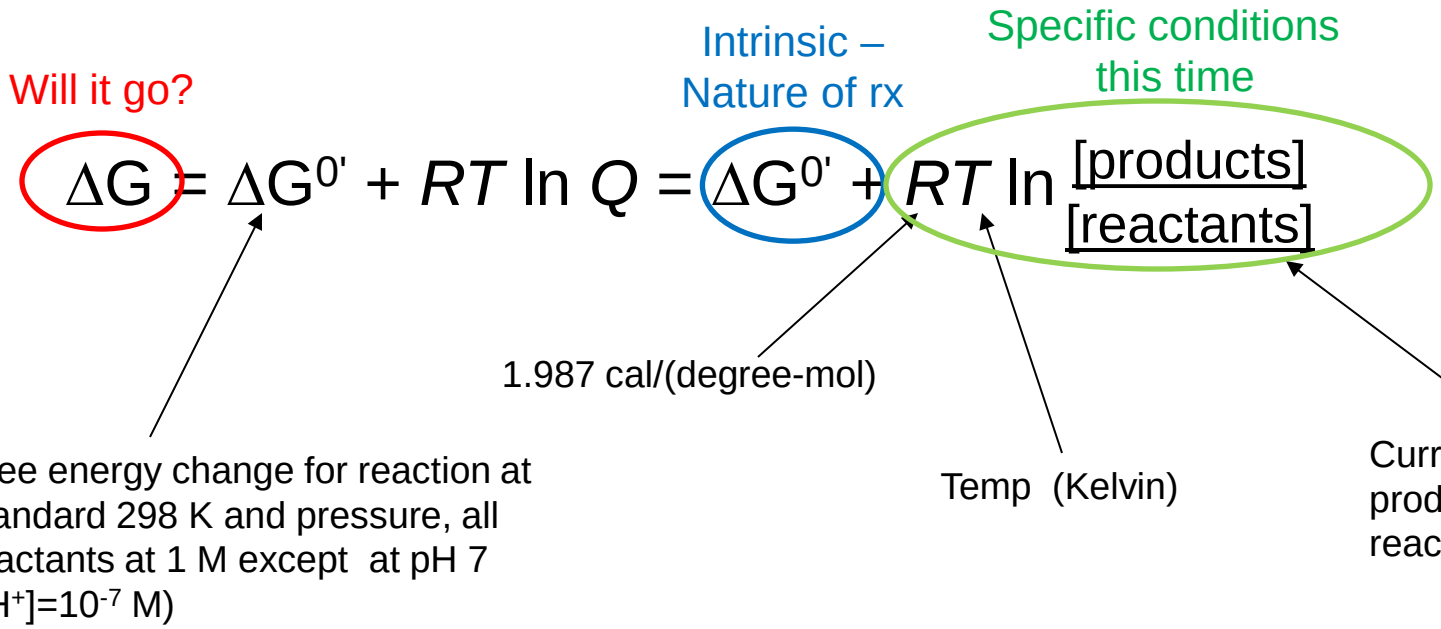


Figure 2-50 *Molecular Biology of the Cell* (© Garland Science 2008)

Gibbs free energy changes predict the direction of chemical and biochemical reactions

$$\Delta G = \Delta H - T\Delta S$$

Free energy is negative if reaction is strongly exothermic ($\Delta H < 0$), or if entropy (ΔS) increases enough to offset endothermic $\Delta H > 0$



Reactions run to equilibrium: level of product at which forward and reverse reactions are balanced

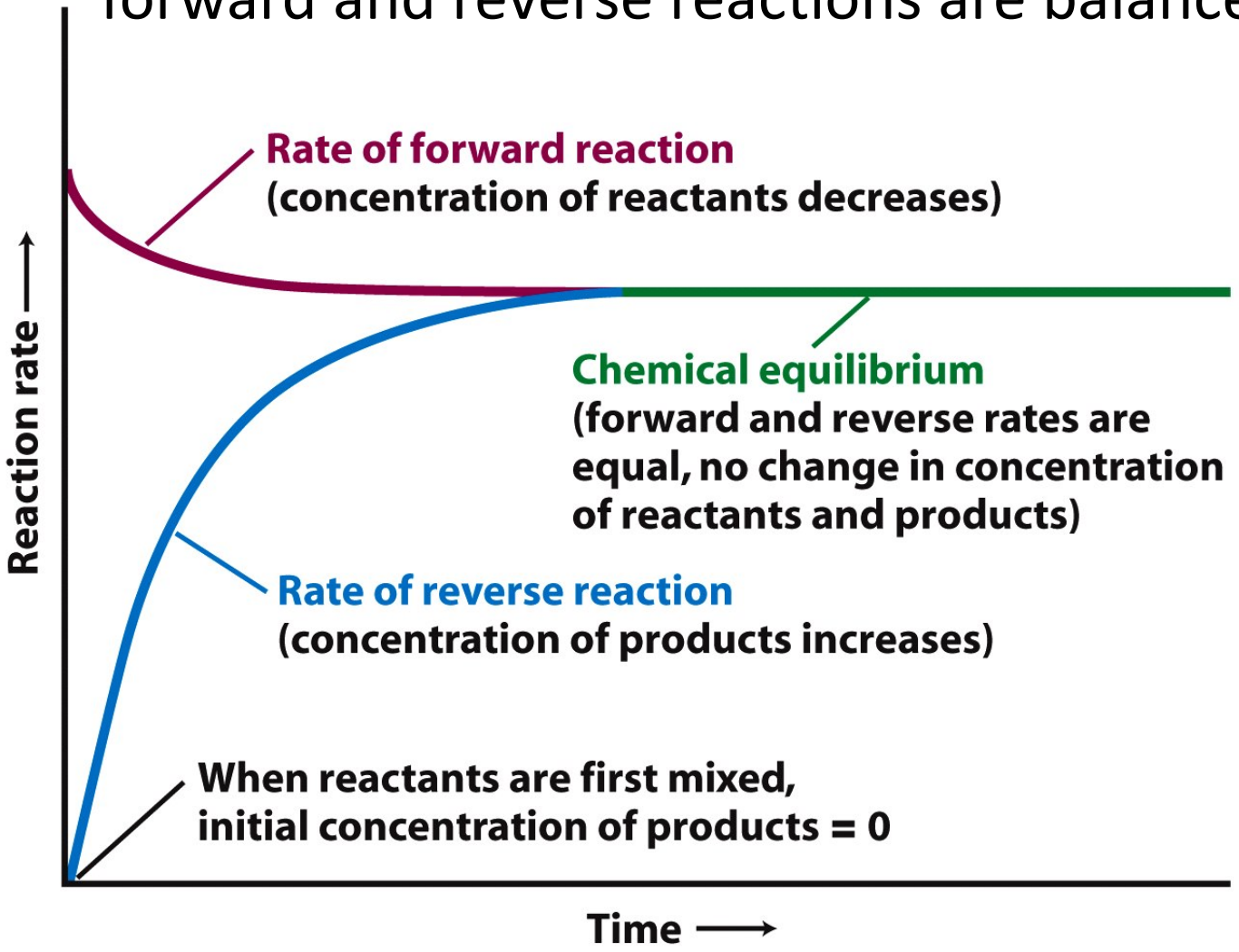


Figure 2-22
Molecular Cell Biology, Sixth Edition

A major type of reaction is binding vs. unbinding of two molecules

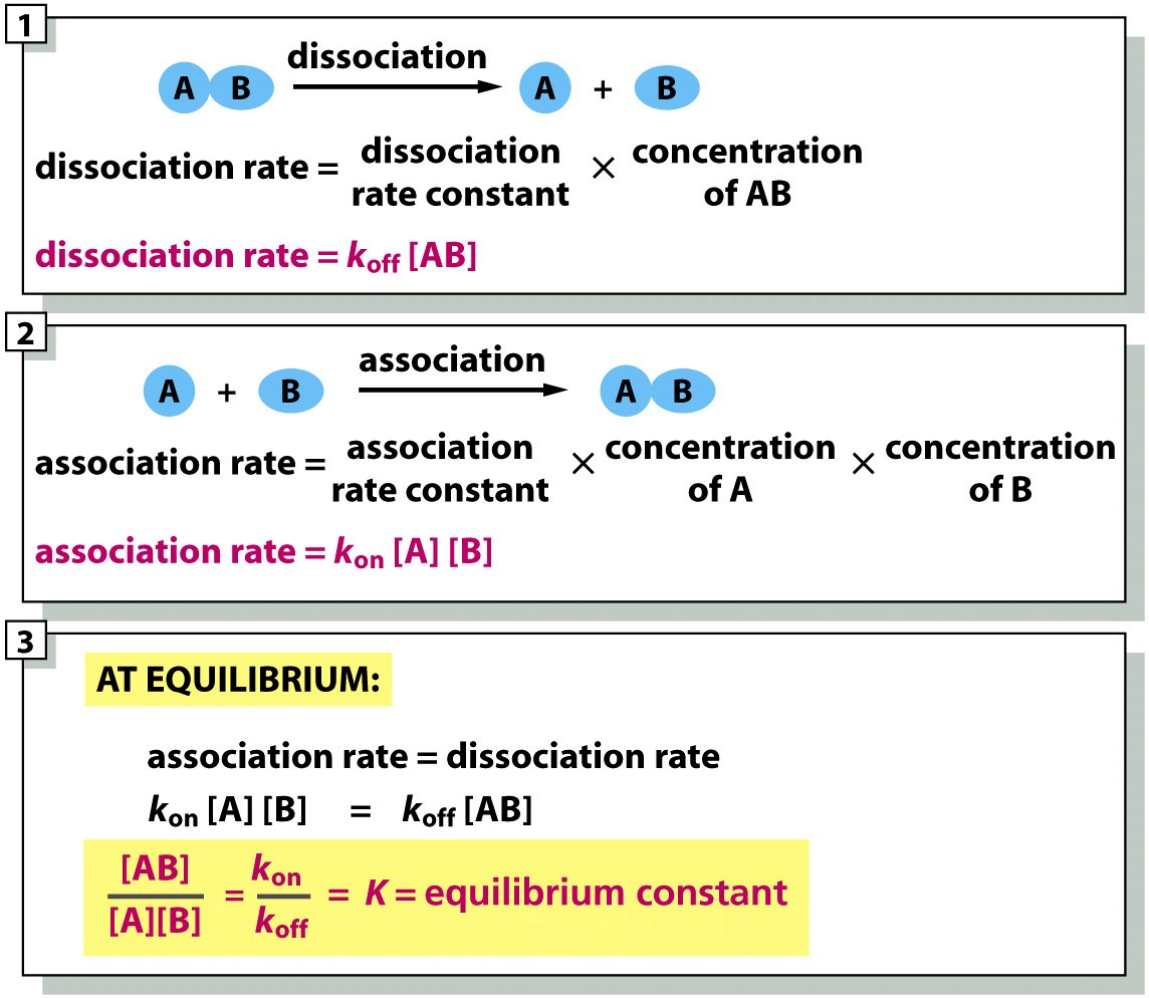


Figure 3-43a *Molecular Biology of the Cell* (© Garland Science 2008)

Equilibrium constant K_{eq} for a reaction depends directly on $\Delta G^{0'}$

At equilibrium, when $\Delta G = 0$,

$\Delta G^{0'} = -RT \ln K_{eq}$

or $K_{eq} = e^{(-\Delta G^{0'}/RT)}$

Or for a binding reaction,

$[AB]/([A][B]) = e^{(-\Delta G^{0'}/RT)}$

If $\Delta G^{0'}$ itself is negative, this makes exponent positive …and [AB] > [A][B]

Figure 3-43c *Molecular Biology of the Cell* (© Garland Science 2008)

The power of a few noncovalent bonds in setting chemical equilibria

Recall: the strength of *ONE* H-bond in water is ~ 1 kcal/mole … a difference of 2-3 H-bonds in a complex can shift equilibrium constant by a factor of 30-100

The relationship between free-energy differences and equilibrium constants (37°C)

equilibrium constant $\frac{[AB]}{[A][B]} = K$ (liters/mole)	free-energy difference of AB minus free energy of A + B (kcal/mole)	free-energy difference of AB minus free energy of A + B (kJ/mole)
1	0	0
10	–1.4	–5.9
10^2	–2.8	–11.9
10^3	–4.3	–17.8
10^4	–5.7	–23.7
10^5	–7.1	–29.7
10^6	–8.5	–35.6
10^7	–9.9	–41.5
10^8	–11.3	–47.4
10^9	–12.8	–53.4
10^{10}	–14.2	–59.4
10^{11}	–15.6	–65.3

Figure 3-43b *Molecular Biology of the Cell* (© Garland Science 2008)

Transcription factor targeting: a quantitative issue

- Transcription factors can bind all DNA somewhat, specific DNA better.
- K_s = specific site binding constant
- K_n= nonspecific site binding constant… includes binding to all other sites in whole genome

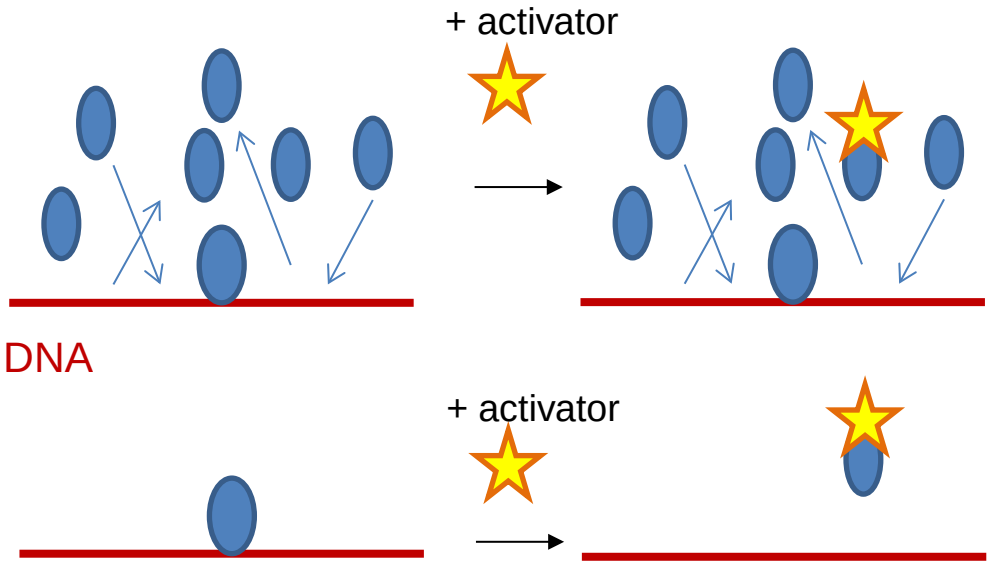
For DNA binding proteins with normal DNA binding affinities, trying to find their sites in a real genome, specificity ***ratio*** Kr is crucial:

- $K_r = K_s/K_n$ ← This is what affects transcription factor potency!

- What matters:
  (1) number of copies of a transcription factor in a cell, (2) size of the genomic "search space" (small for bacteria, large for us), and (3) the factor's "preference ratio" for specific site, or K_r
- You get half max occupancy of a specific site when TF concentration is enough so that:
  [TF concentration] ~ [Concentration of nonspecific sites]/K_r

Activator TFs can work intermittently (bacteria or eukaryotes)… but Bacterial repressors need to block polymerase loading all the time

- High requirement for specificity in bacterial repressors
- Loading up high numbers of repressor molecules cannot do the trick, if regulation is to be sensitive to signals…



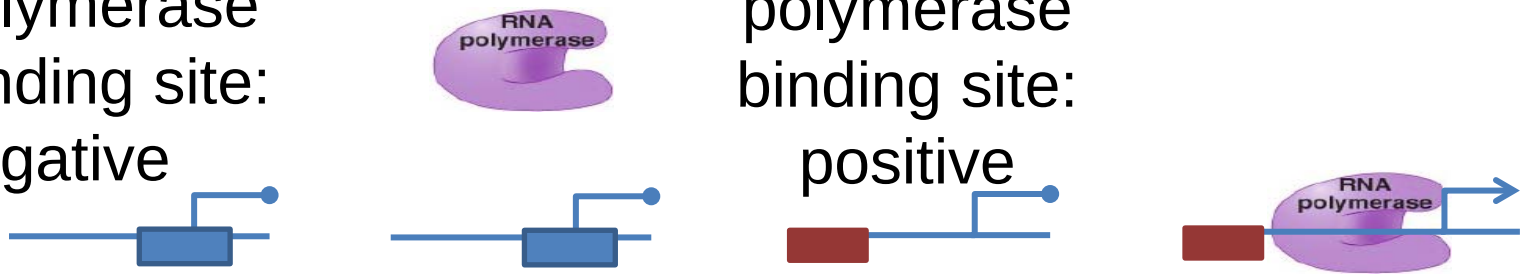
How can you get stronger effects with weakly specific factors and modest copy numbers?

- Cooperative action
- Cooperative binding!
  - We will talk about this a lot in context of eukaryotic gene regulation
- Gene networks
  - Small circuit elements
  - Larger networks that drive developmental processes & choices

Using transcriptional regulation for logic operations

LacI site ***crosses*** polymerase binding site: negative

CAP (CRP) site ***adjoins*** polymerase binding site: positive



In bacterial systems, underlying DNA sequence with orientation of binding sites crucially specifies the roles TFs can play… when they occupy sites

Length of DNA available for regulation is a constraint for transcription factor action

Bacterial promoter: a very crowded place

- Promoter loading
- Repressor obstruction of polymerase
- Surfaces for positive regulators to touch polymerase complex & σ factors

Length of DNA available for regulation is a constraint for transcription factor action

Can one use sequence beyond the promoter?

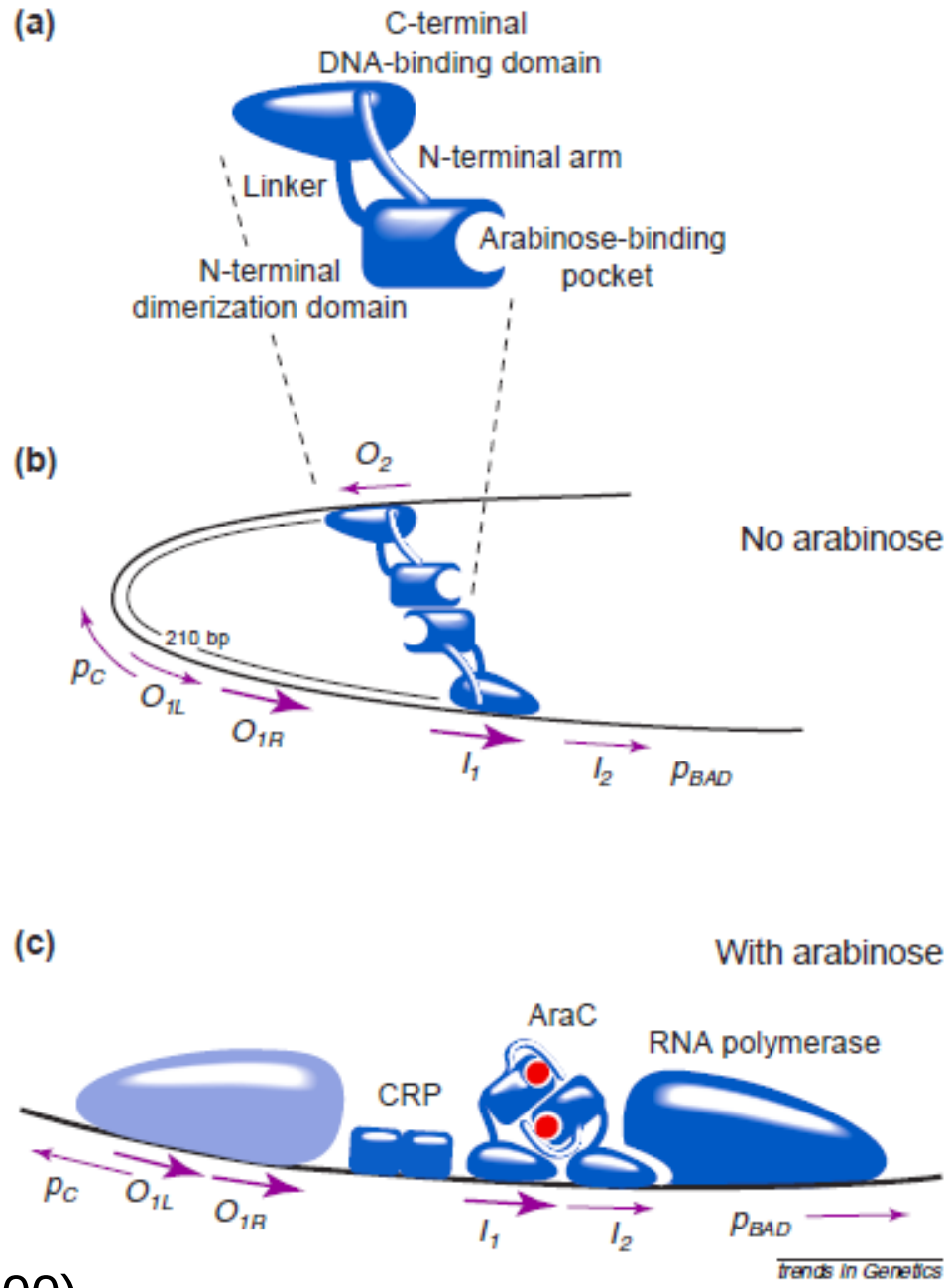
- Can transcription factor binding away from promoter be looped to affect loading at the promoter?

Yes: and the length itself can act to make sure that there is specificity in the complexes that will work

- For single-stranded nucleic acids, even 12 nt can form a hairpin
- For double-stranded DNA, helix is stiff – foldback distance ("persistence length") is about 500 bp
- Interactions within ~200 bp distances or less can only loop DNA in a geometrically specific way, pulled by proteins

Short time scales, short DNA lengths, high precision of bacterial interactions

The Arabinose operon:
A bifunctional TF that loops DNA when it blocks transcription (no Ara) and releases loop when it activates (+ Ara)

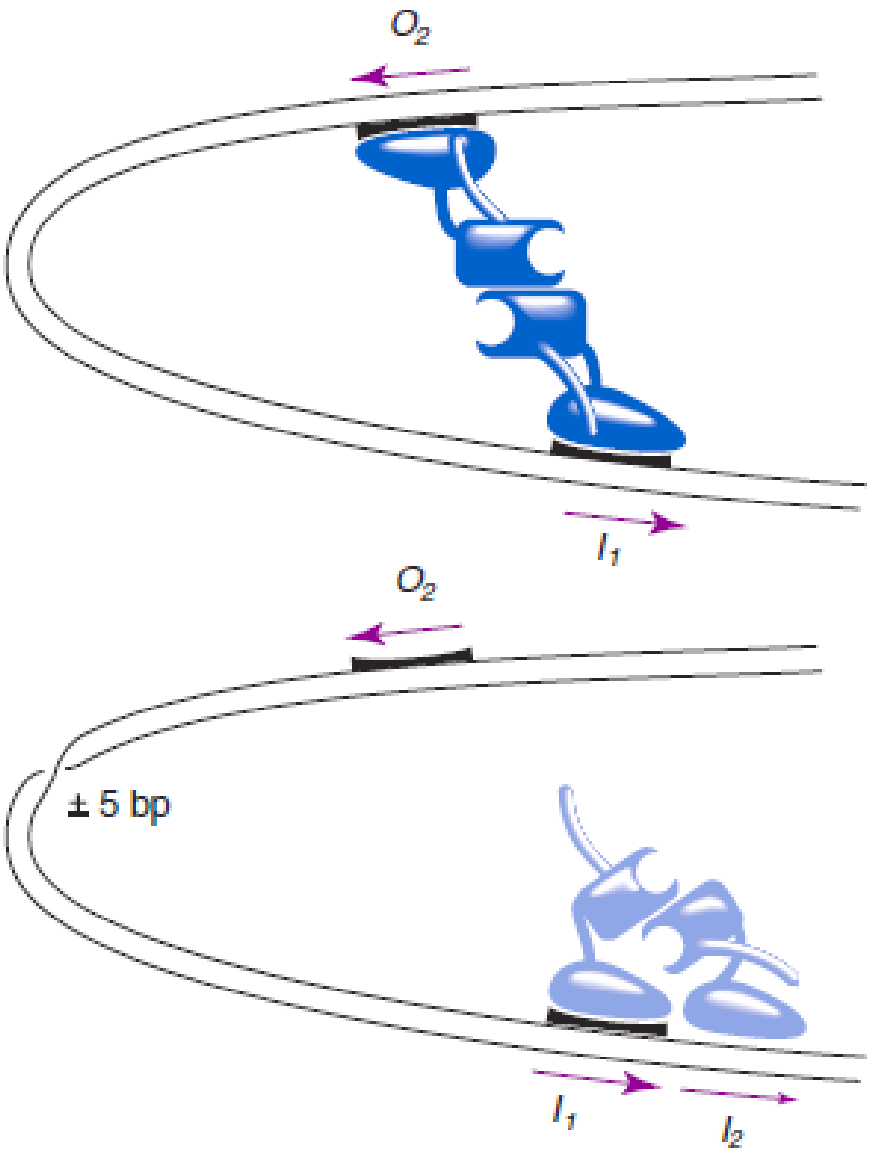


(R. Schleif, 2000)

Even when bacterial TFs loop DNA, they do so under extreme constraint due to short distances…

Adding half-integral DNA turns between the O2 and I1 half-sites interfered with repression of pBAD in the absence of arabinose, causing a 5-10x increased basal level of transcription. Adding integral numbers of turns did not interfere with this repression.

FIGURE 4. The helical-twist experiment that demonstrated DNA looping



(R. Schleif, 2000)

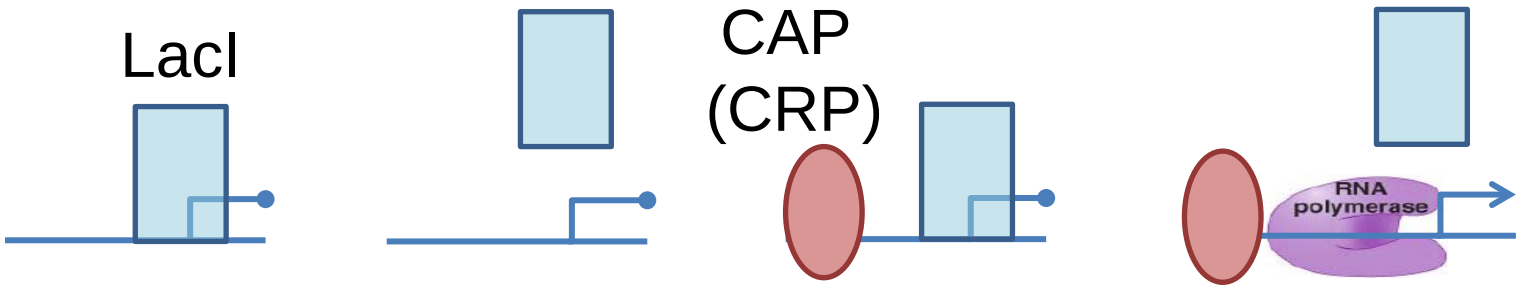
The breakthrough of multicellular eukaryotes: figuring out a way to make more extended DNA available for regulation

- Vastly increased the number of possibilities for regulatory input

- Made possible true “or” logic at a single promoter

We will talk about this a lot more: a key point

Using transcriptional regulation for logic operations: more than one input



If No lactose
And glucose

If lactose
And glucose

If no lactose
And no glucose

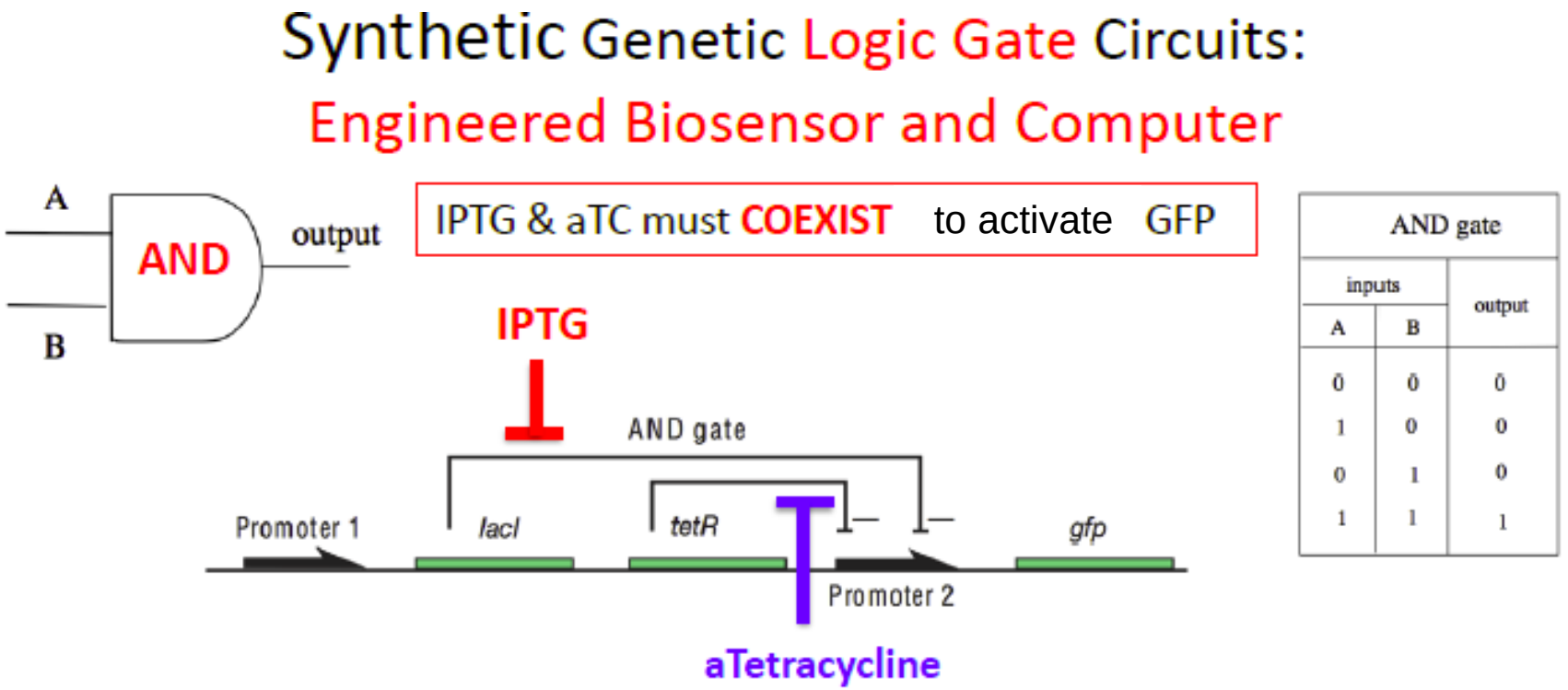
If lactose
And no glucose

Lac operon:

a series of if… and if…then statements

mediated by discrete TF: binding site interactions and defined TF roles

Using these principles, one can build one's own synthetic logic circuits



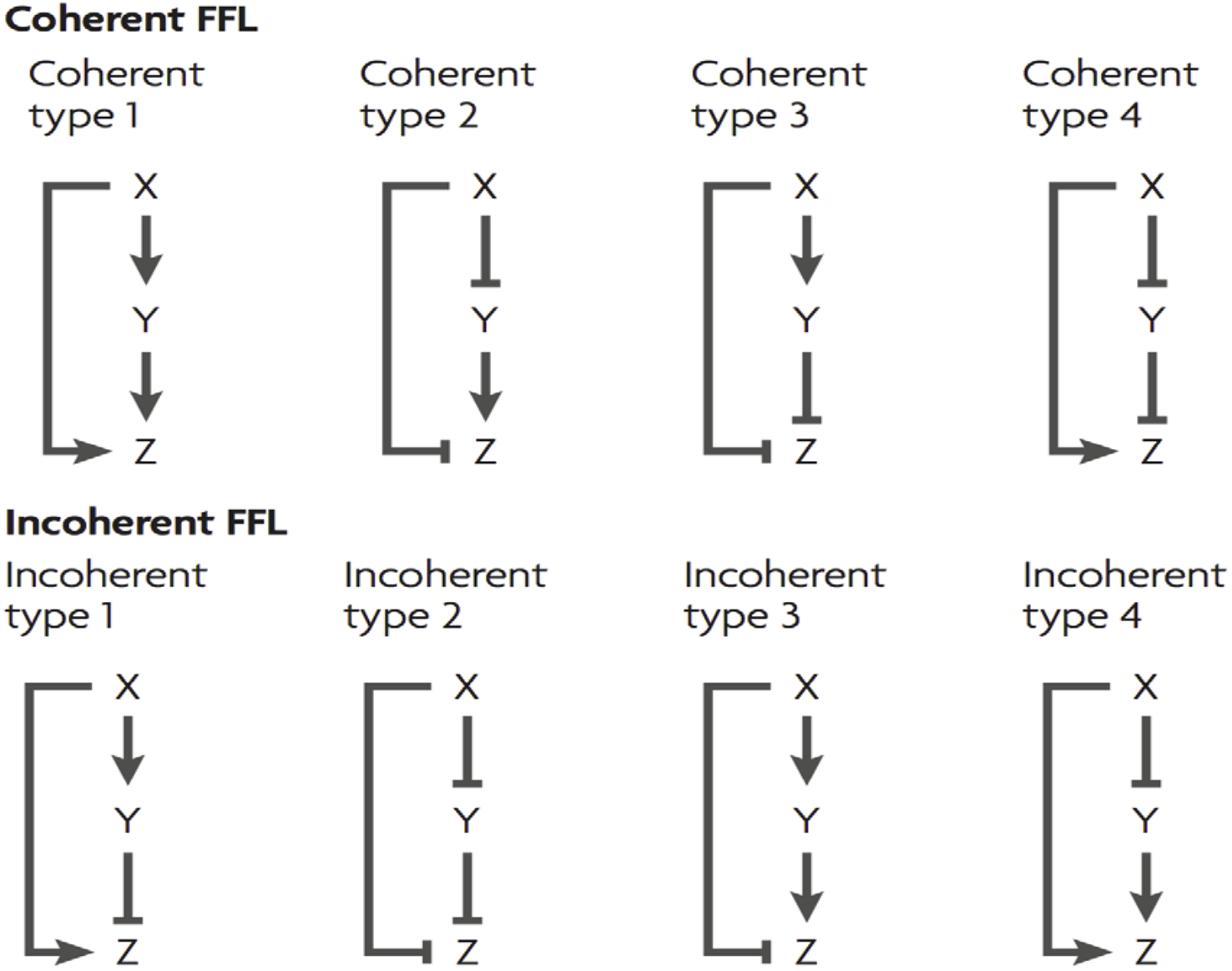
AND gate		
inputs		output
A	B	
0	0	0
1	0	0
0	1	0
1	1	1

IPTG: a Lac inducer
aTetracycline: a Tet operon inducer
Promoter 2: blocked by sites for tetR and lacI

Slide courtesy of Xun Wang

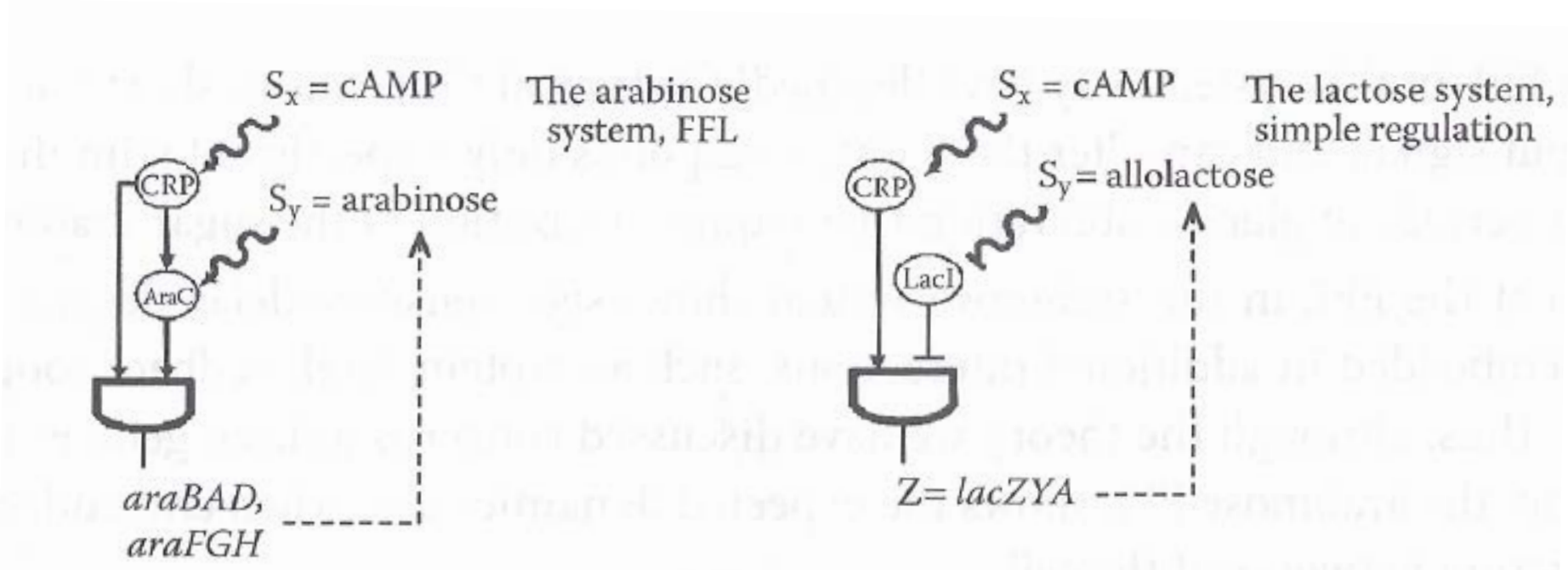
Hasty, McMillen & Collins 2002

Genes can be linked by regulation into network circuits that perform specific biological roles: dynamic signal processing



U. Alon, 2007

Being able to combine two positive inputs to turn on downstream genes can modulate kinetics of response

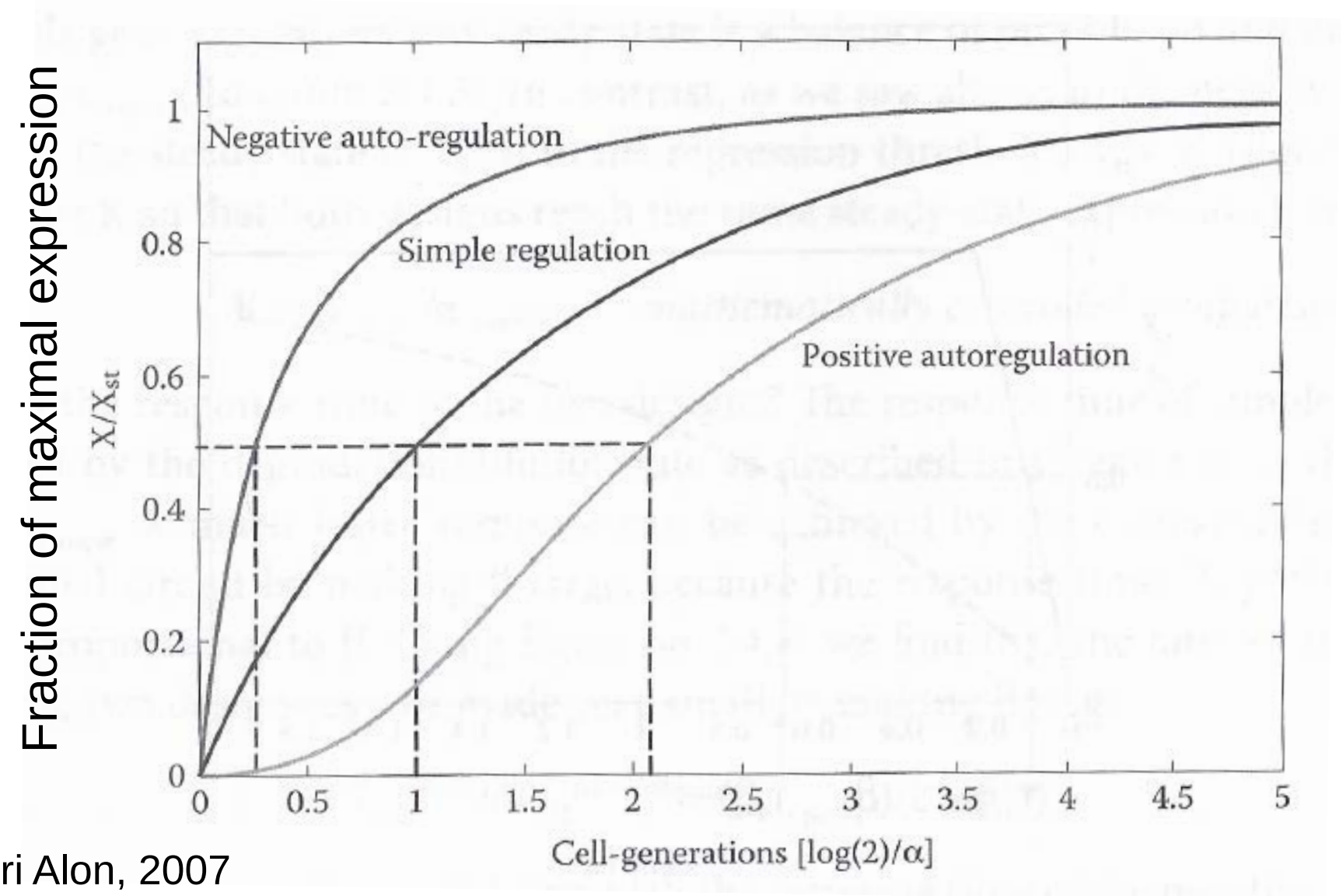


"Feed forward" because one positive input depends on other

"Simple" because positive and negative inputs are independent

Uri Alon, 2007

Kinetics of gene activation can depend on gene circuit feedbacks



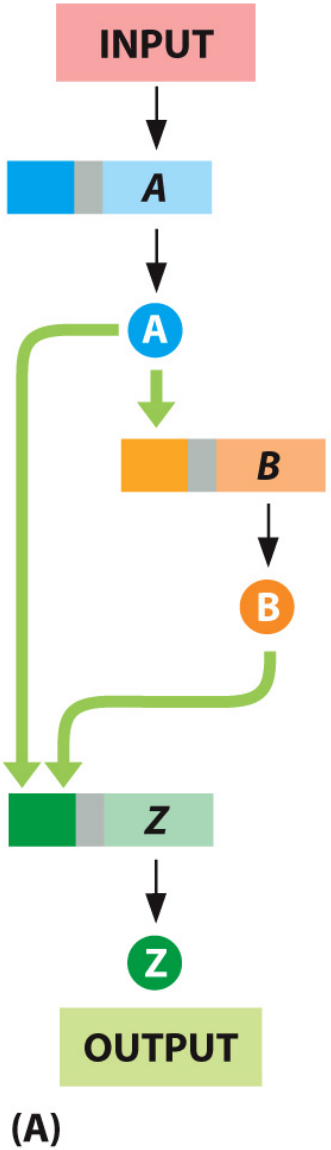
Uri Alon, 2007

Regulation can affect transcription factors' expression *or* transcription factors' activities

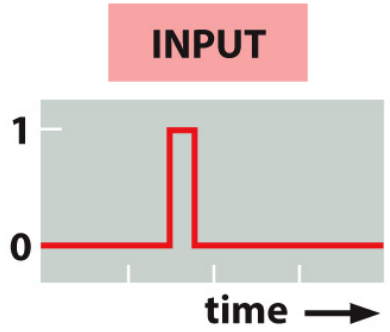
- Expression of a TF establishes potential for its target genes to receive its input, but does not ***dictate*** output

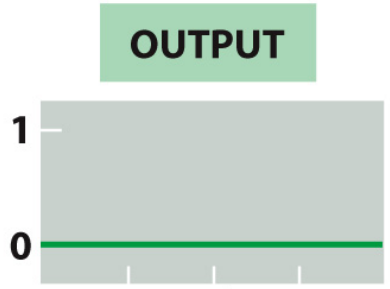
- Ways to modulate TF activity
  - Prevent/enable binding to DNA by interaction with small molecule (almost exclusively in bacteria)
  - Enable ability to recruit polymerase by enzymatic modification (e.g. kinase, triggered by signaling)
  - Switch function between activation and repression by enzymatic modification (mostly in eukaryotes)
  - Prevent/enable binding to DNA by cooperation with other bound TFs

Transcription circuits allow cells to carry out logic operations on temporally varying inputs

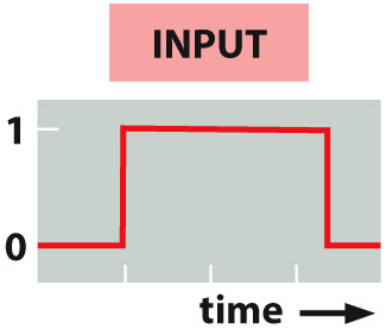


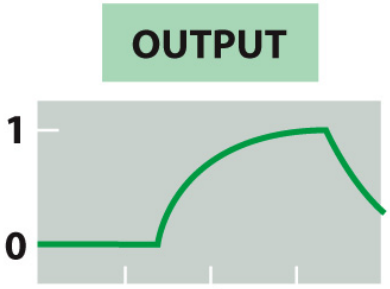
(A)





(B)





(C)

Feed forward loop acting as a filter to ignore transient "noise"

Figure 7-41 Molecular Biology of the Cell 6e (© Garland Science 2015)

Autoregulation and cross-regulation of genes encoding transcription factors create circuits with different properties

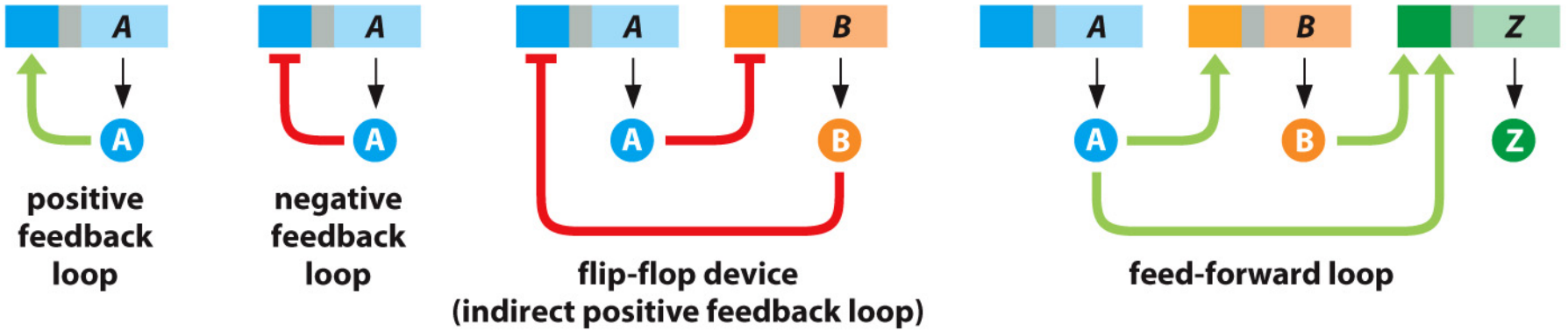


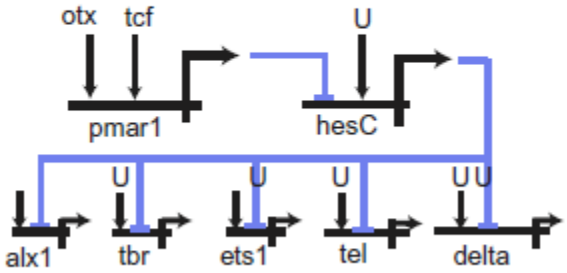
Figure 7-40 Molecular Biology of the Cell 6e (© Garland Science 2015)

Canonical circuit elements are also used in multicellular organism development to make one part of embryo different from another

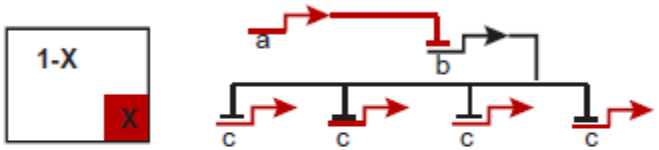
A Job: X, 1-X spatial specification function

A1 Subcircuit: double negative gate

From GRN:



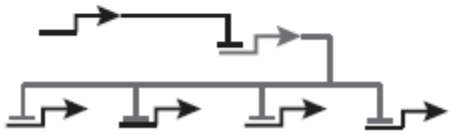
a: Repressor1
b: Repressor2
c: Target genes



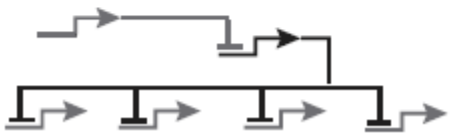
Subcircuit output

	a	b	c
X	1	0	1
1-X	0	1	0

In X



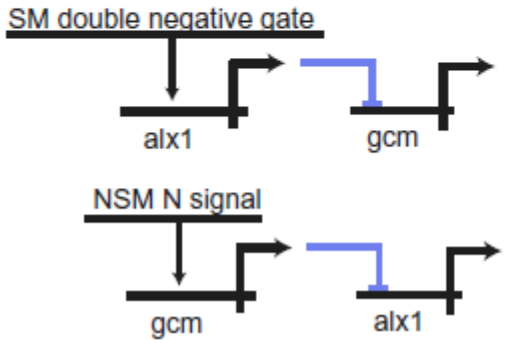
In 1-X



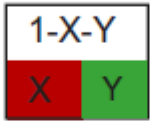
C Job: mutual exclusion of spatial regulatory state

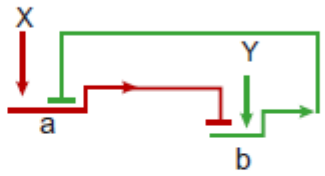
Subcircuit: reciprocal repressor

From GRN:



a: (X) specification gene: Rep
b: (Y) specification gene: Rep

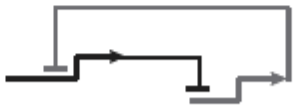




Subcircuit output

	a	b
X	1	0
Y	0	1
1-X-Y	0	0

In X



In Y

Peter & Davidson, 2009

When do you need transcriptional regulation complexity?

- For single-celled organisms, for tuning sensitivity & kinetics of responses to environment
- Also for establishing major biological state changes (sporulation in *Bacillus*)

- For multicellular organisms – ALWAYS
  - Making one cell different from another
  - Modulating responses not only to cell's own needs but also to current physiology of whole organism