Network motifs in the transcriptional regulation network (of Escherichia coli):

Janne.Ravantti@Helsinki.Fi

Computational Systems Biology seminar, 28.03.2007

Contents:

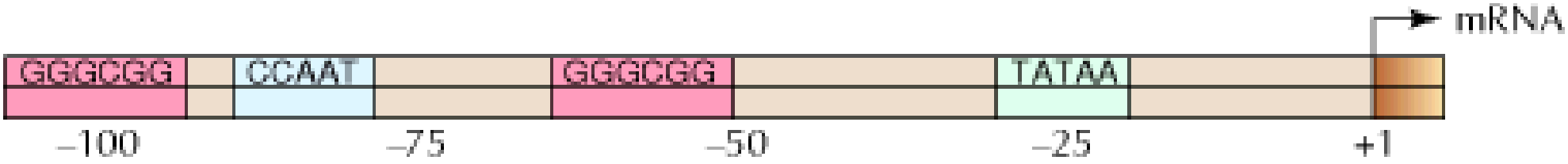
- Transcription Networks (aka. “The Very Boring Biology Part”)
- Network Motifs
- Examples of Network Motifs from E.Coli
- Critique / Problems
- Conclusions

Main References:

- Shen-Orr SS, Milo R, Mangan S, Alon U.: Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002 May;31(1):64-8.
- An Introduction to Systems Biology: Design Principles of Biological Circuits. Uri Alon, 2006, ISBN: 9781584886426
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U.: Network motifs: simple building blocks of complex networks. Science. 2002 Oct 25;298(5594):824-7.

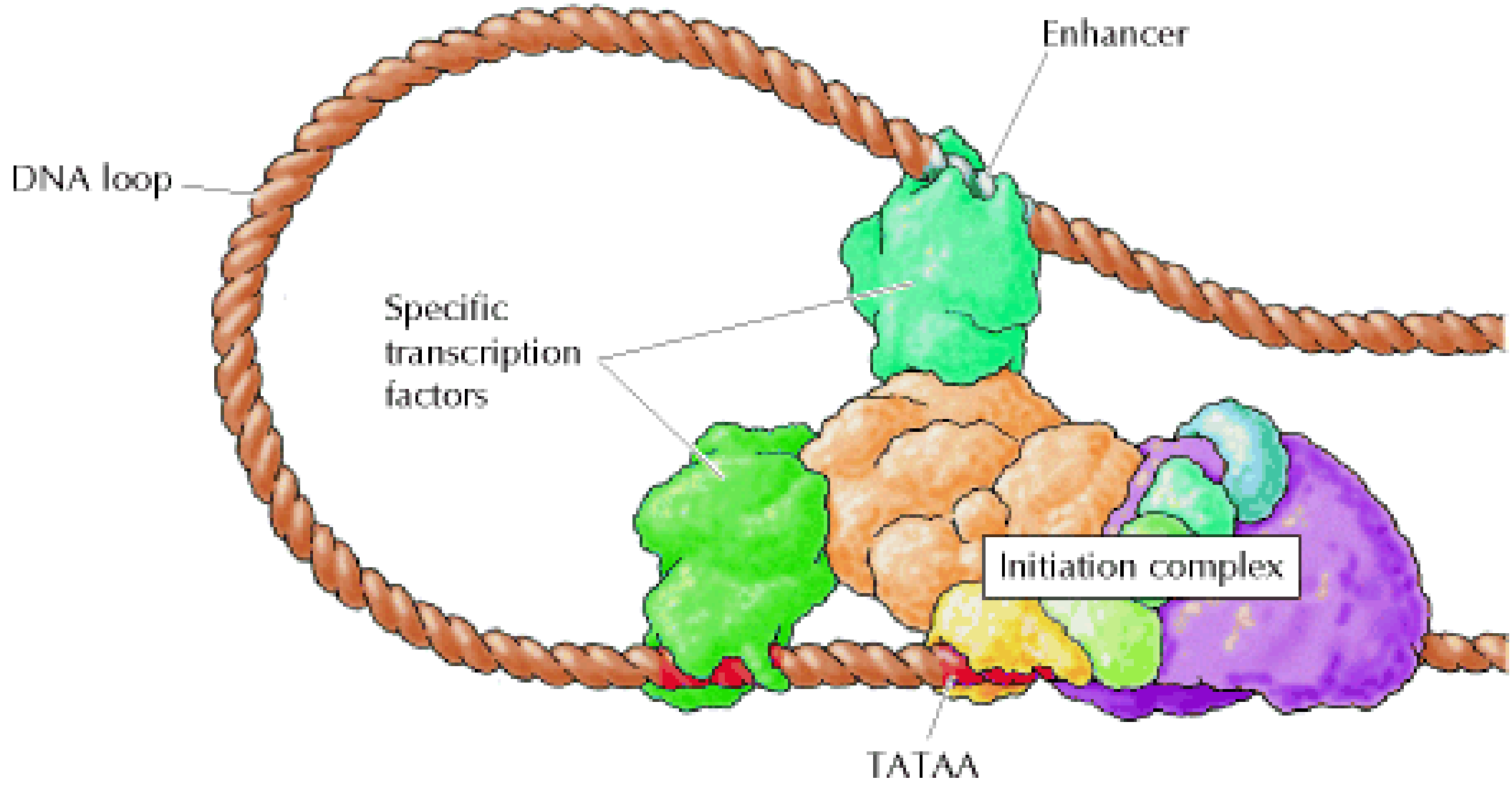
Transcription (1/3):

- Transcription: The synthesis of an RNA molecule from a DNA template ("reading a gene/protein from the piece of DNA")
- Transcription is controlled by proteins ("transcription factors") that bind to specific regulatory sequences



- E.g. 3 elements upstream of the TATA box required for efficient transcription

Transcription (2/3):



Transcription (3/3):

- Interaction between transcription factors and genes can be described by a transcription network:
  - Nodes are genes
  - Edges represent how genes regulate other genes (or themselves!) => directed graph
- Transcription factors can:
  - increase transcription rate (“activatiors”)
  - decrease -:- -:- (“repressors”)

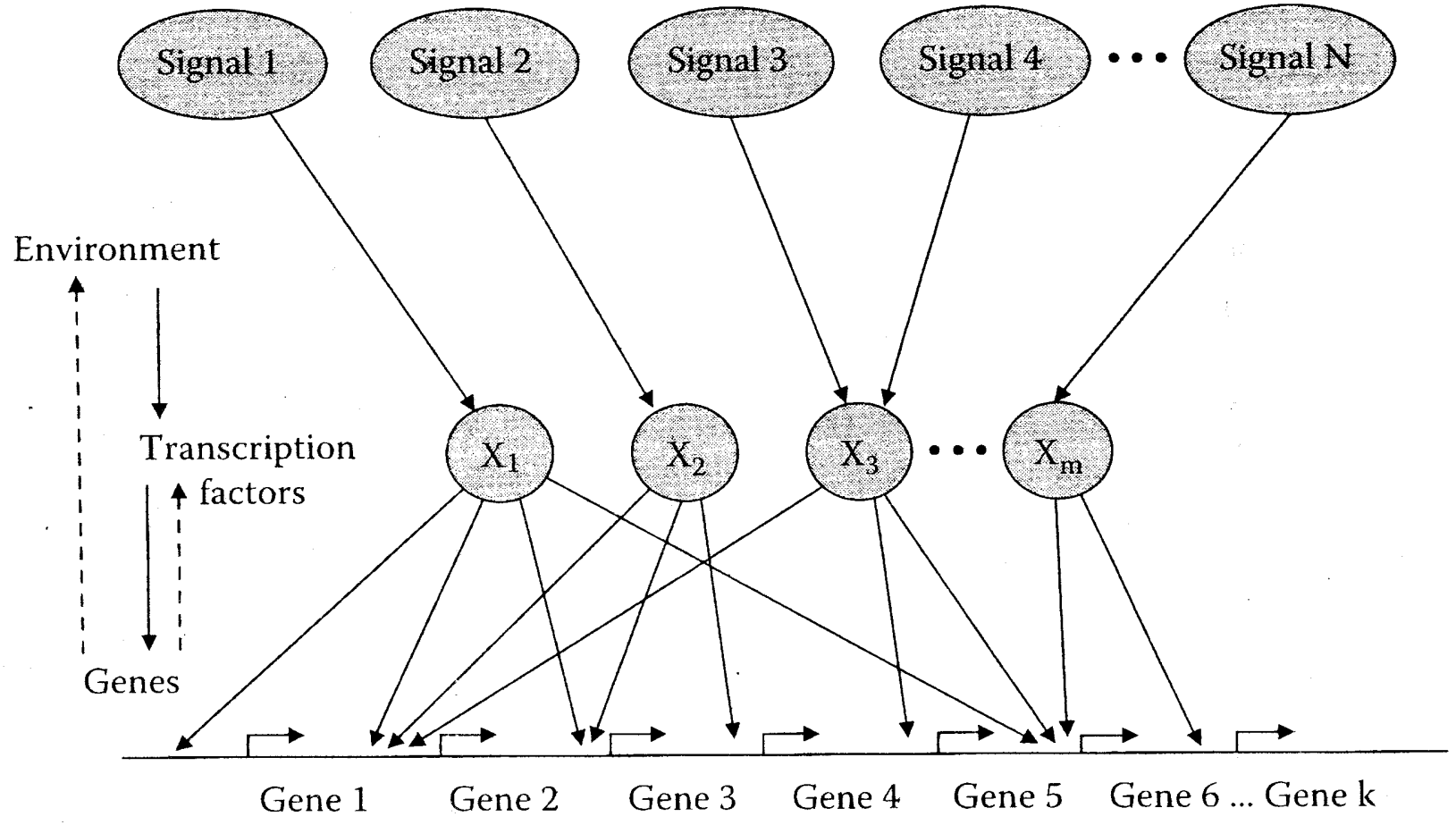
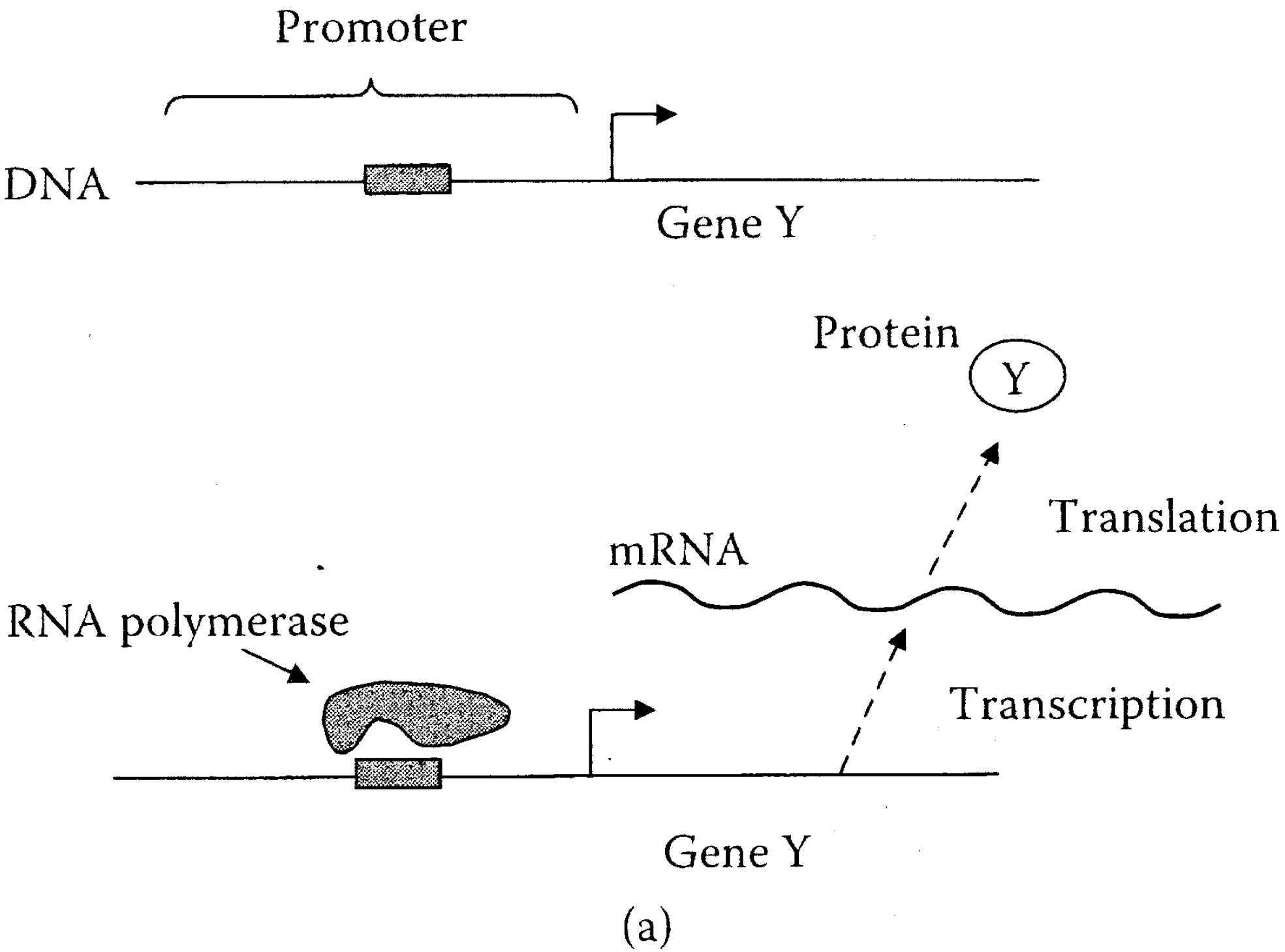
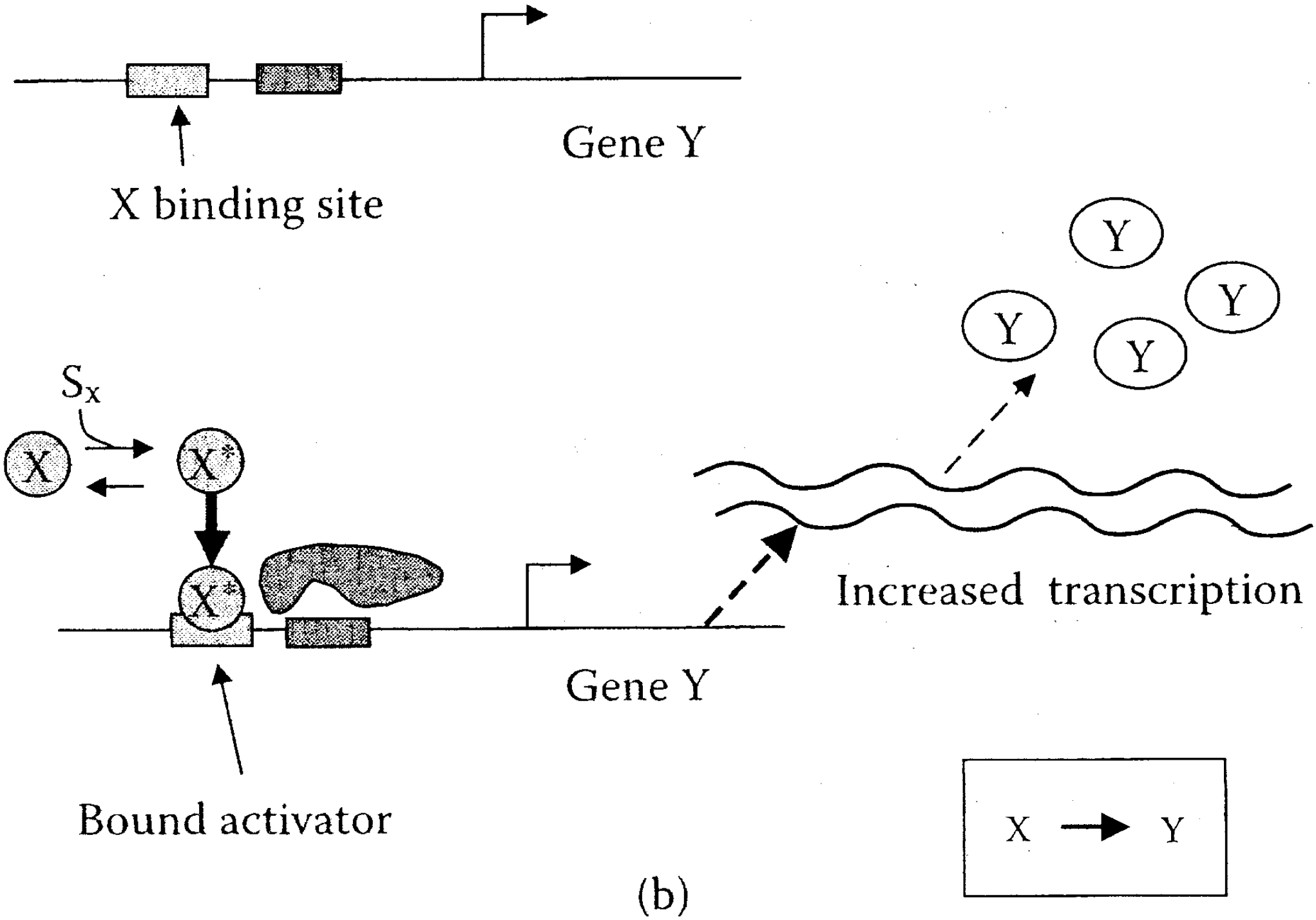
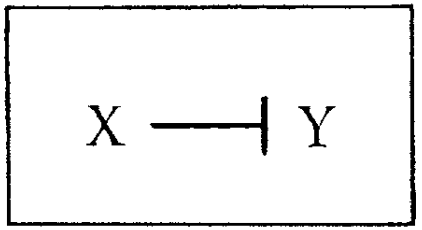
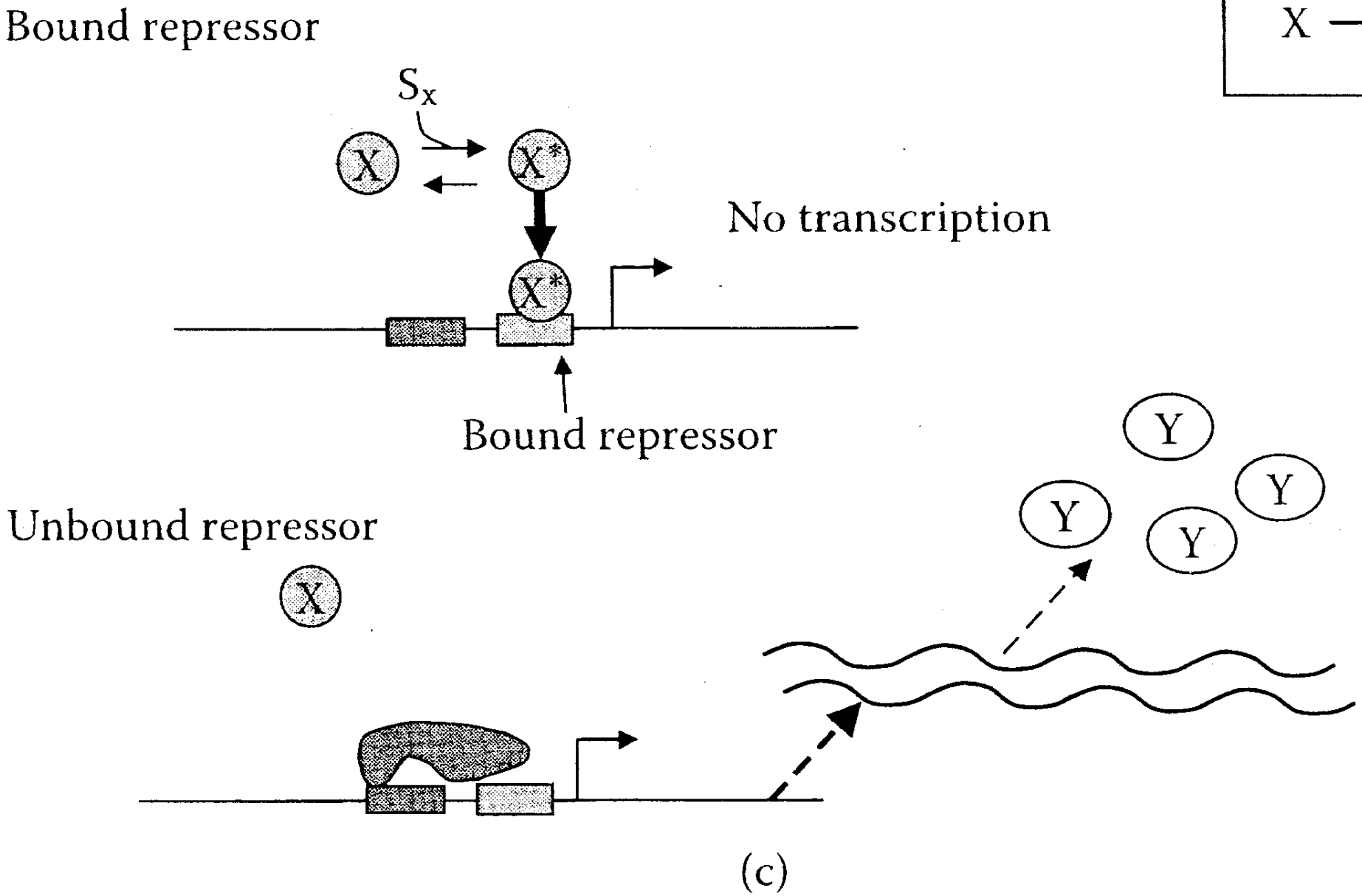


FIGURE 2.1 The mapping between environmental signals, transcription factors inside the cell, and the genes that they regulate. The environmental signals activate specific transcription factor proteins. The transcription factors, when active, bind DNA to change the transcription rate of specific target genes, the rate at which mRNA is produced. The mRNA is then translated into protein. Hence, transcription factors regulate the rate at which the proteins encoded by the genes are produced. These proteins affect the environment (internal and external). Some proteins are themselves transcription factors that can activate or repress other genes.


Promoter
DNA
Gene Y
Protein
Y
Translation
mRNA
RNA polymerase
Transcription
Gene Y
(a)


Gene Y
X binding site
Y
Y
Y
Y
S_X
X
X*
X*
Increased transcription
Gene Y
Bound activator
X → Y
(b)


X
Y


Bound repressor
S_X
X
X*
No transcription
X*
Bound repressor
Unbound repressor
X
Y
Y
Y
Y

(c)

Timescales:

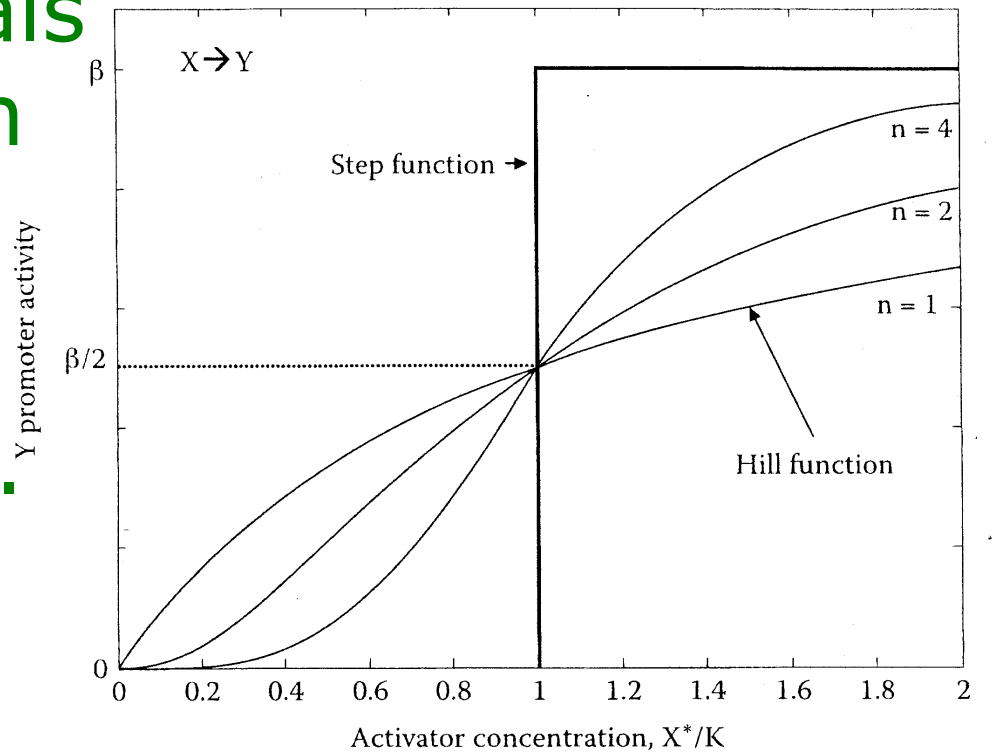
TABLE 2.2 Timescales for the Reactions in the Transcription Network of the Bacterium *E. coli* (Order of Magnitude)

Binding of a small molecule (a signal) to a transcription factor, causing a change in transcription factor activity	~1 msec
Binding of active transcription factor to its DNA site	~1 sec
Transcription + translation of the gene	~5 min
Timescale for 50% change in concentration of the translated protein (stable proteins)	~1 h (one cell generation)

=> Transcription factor activity ≈ steady state on timescale of changes in protein levels

Network dynamics:

- Separation of different timescales simplifies description of network dynamics
- Real gene input signals can be described with S-shaped ("Hill function"):
- More in Alon's book... (today's topic: network motifs)



Autoregulation:

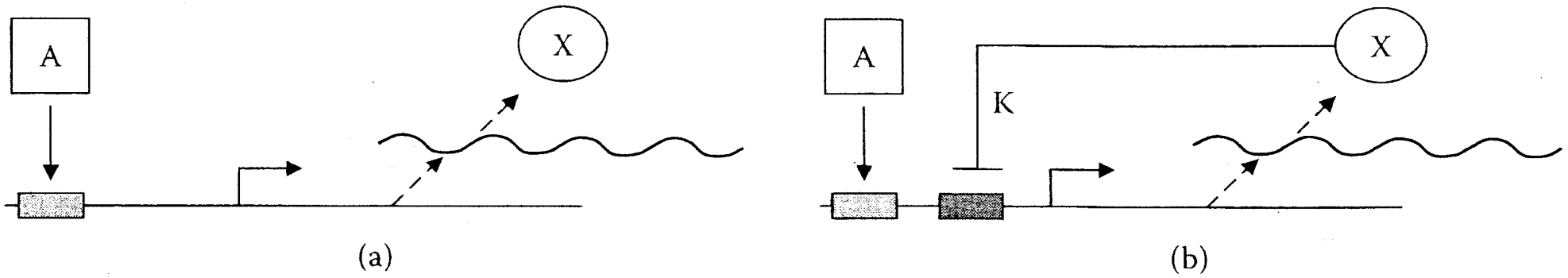
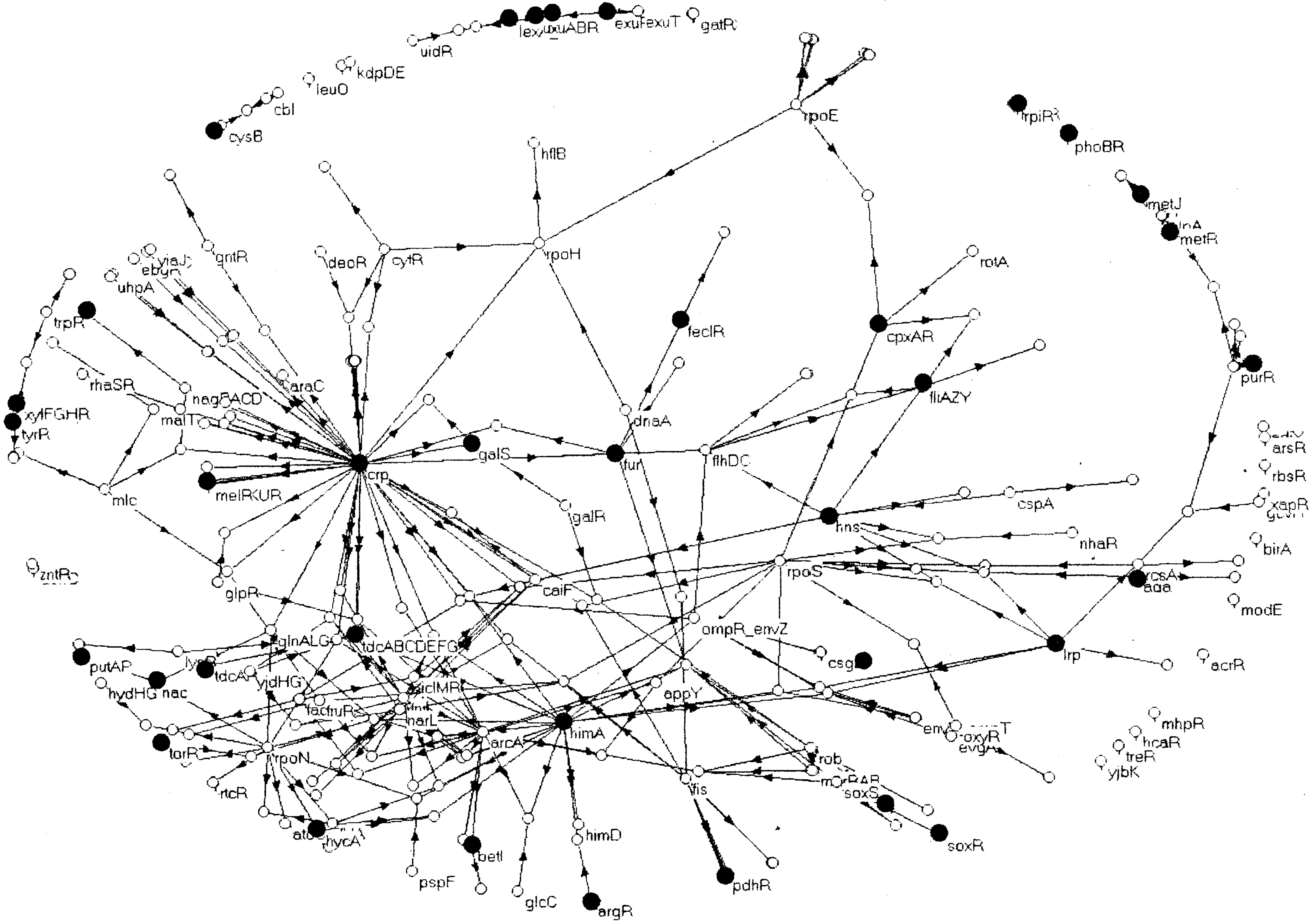


FIGURE 3.2 Simple regulation and negative autoregulation. (a) Gene X is simply regulated by A. (b) A gene X that is negatively autoregulated; that is, it is repressed by its own gene product, the repressor X. The gene is also simply regulated by A. Repressor X binds a site in its own promoter and thus acts to repress its own transcription. The symbol ⊣ stands for repression. The repression threshold is K, defined as the concentration of X needed to repress the promoter activity by 50%.


exuTexuT
gatR
uidR
kdpDE
leuO
cbl
cysB
hflB
rpoE
rpiR
phoBR
metJ
metR
gntR
deoR
cytR
rpoH
rotA
ebgR
uhpA
trpR
feclR
cpxAR
purR
rhaSR
nagBACD
araC
fliAZY
xylFGHR
malT
dnaA
tyrR
crp
galS
fur
flhDC
arsR
rbsR
melRKUR
galR
cspA
hns
mlc
rhaR
birA
rpoS
caiF
glpR
modE
ompR_envZ
glnALG
tdcABCDEFG
csg
lrp
acrR
putAP
tdcAR
yjdHG
hydHG
nac
appY
fadfruR
narL
arcA
himA
mhpR
hcaR
treR
yjbK
torR
rpoN
rob
rtcR
fis
soxS
himD
betl
pspF
soxR
pdhR
glcC
argR

Network motifs in the transcriptional regulation network of E. coli (1/4):

- From: Orr et.al; Nat.Genetics (2002)
- Main idea: systemically break down networks into basic building blocks and analyze the blocks

  => find out the wiring of the cell

Continues... (2/4):

- Network motifs: patterns of interconnections that recur in many different parts of a network at frequencies much higher than those found in randomized networks
- Main result: network is composed of repeated appearances of three highly significant motifs

Continues... (3/4):

- Each network motif has a specific function in determining gene expression, such as generating temporal expression programs (control) and governing the responses to fluctuating external signals (filtering)
- May help define the basic computational elements of other biological networks

Continues... (4/4):

- Data consists of established interactions in which a transcription factor directly binds a regulatory site
- Only operon (== bunch of genes that are transcribed together) level of regulation considered
- Number of operons: 424; number of interactions: 519

(≈ 4300 protein coding genes known...)

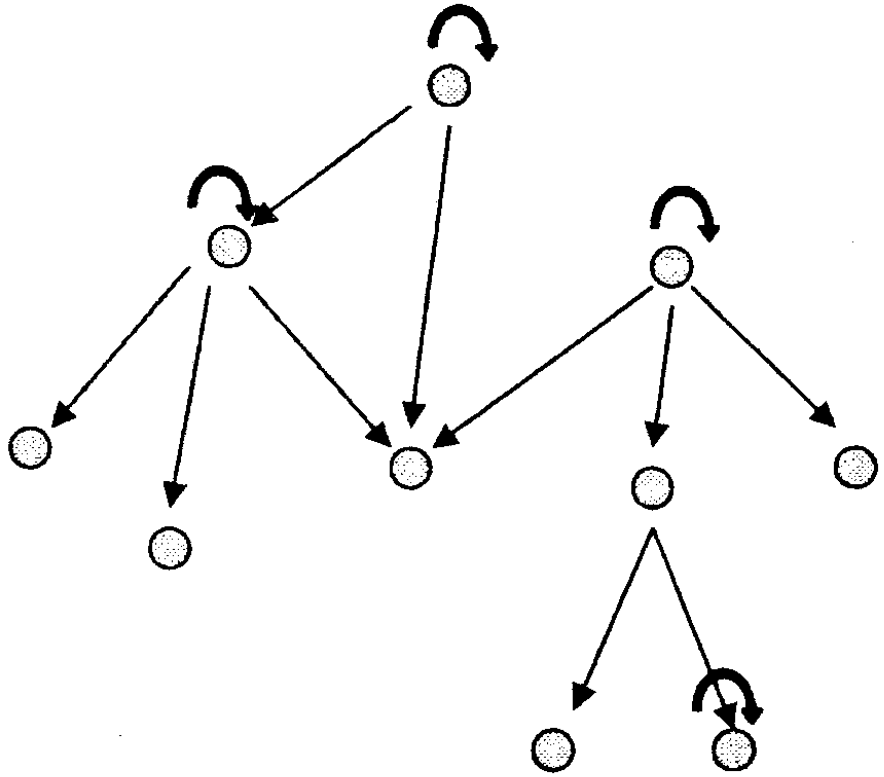
Network Motifs:

- Network motifs are overrepresented subgraphs
- may represent components of functional modules
- Null-hypothesis: ensemble of random networks with same characteristics as the real network

Generation of randomized networks (1/2):

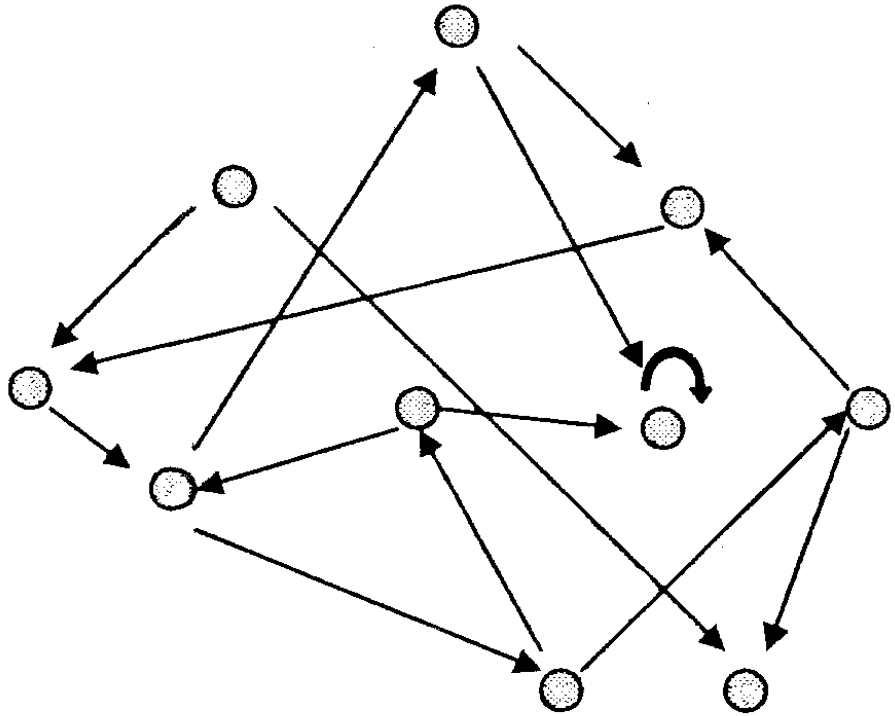
- The real network is compared to an esemble of random networks
- The number of nodes and edges are kept same as in real network (E.Coli: n=424, e=519)
- The total number of generated random networks was 1000 per experiment

'Real' network



N = 10 nodes
E = 14 edges
N_{self} = 4 self-edges

Randomized network
(Erdos – Renyi)



N = 10 nodes
E = 14 edges
N_{self} = 1 self-edge

Generation of randomized networks (2/2):

- Markov-chain:
  - Randomly select a pair of directed edges A -> B and C -> D
  - The two edges are then rewired, so A connects to D, while C connects to B
  - If one or both of these new links already exist, the step is aborted and a new pair of edges is selected. (prevents the appearance of multiple edges connecting the same pair of nodes)
  - Repeat until random enough...

Control of (n-1)-motifs:

- Each randomized network esemble has the same (n-1)-node subgraph count as the real network

  => avoids high significance of n-node graph that actually only has significant (n-1)-subnode
- Implemented as Metropolis Monte-Carlo simulation for networks where n>=4

Network motif detection:

- Essentially, the problem is to find and count all n-node subgraphs from each network (real & randomized)
- Original paper uses recursive counting of connections until n-node subgraph is reached
- Later papers use subgraph isomorphism algorithm (NP-complete) to solve the problem

Selection of a network motif:

1) Probability that the motif appears in randomized network more often or equally than in real network is < 0.01

2) Motif appears in the real network in distinct set of nodes more than 4 times

3) Appears in real network significantly more often than in randomized networks $N_{real} - N_{rand} > 0.1 * N_{rand}$

=> Measure of significance: $Z = (N_{real} - N_{rand})/SD$

Results:

Table 1 • Statistics of occurrence of various structures in the real and randomized networks

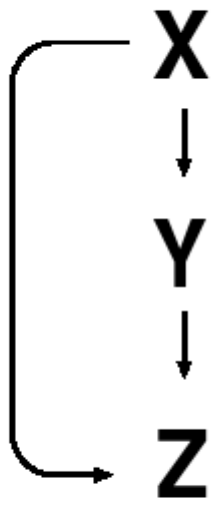
Structure	Appearances in real network	Appearances in randomized network (mean ± s.d.)	*P* value
Coherent feedforward loop	34	4.4 ± 3	$P < 0.001$
Incoherent feedforward loop	6	2.5 ± 2	$P \sim 0.03$
Operons controlled by SIM (>13 operons)	68	28 ± 7	$P < 0.01$
Pairs of operons regulated by same two transcription factors	203	57 ± 14	$P < 0.001$
Nodes that participate in cycles*	0	0.18 ± 0.6	$P \sim 0.8$

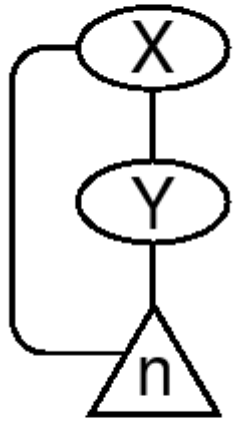
*Cycles include all loops greater than size 1 (autoregulation). *P* value for cycles is the probability of networks with no loops.

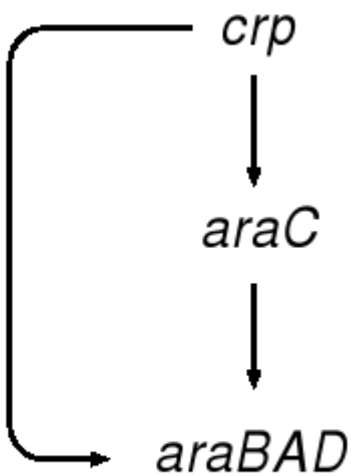
Feedforward loops (1/3):

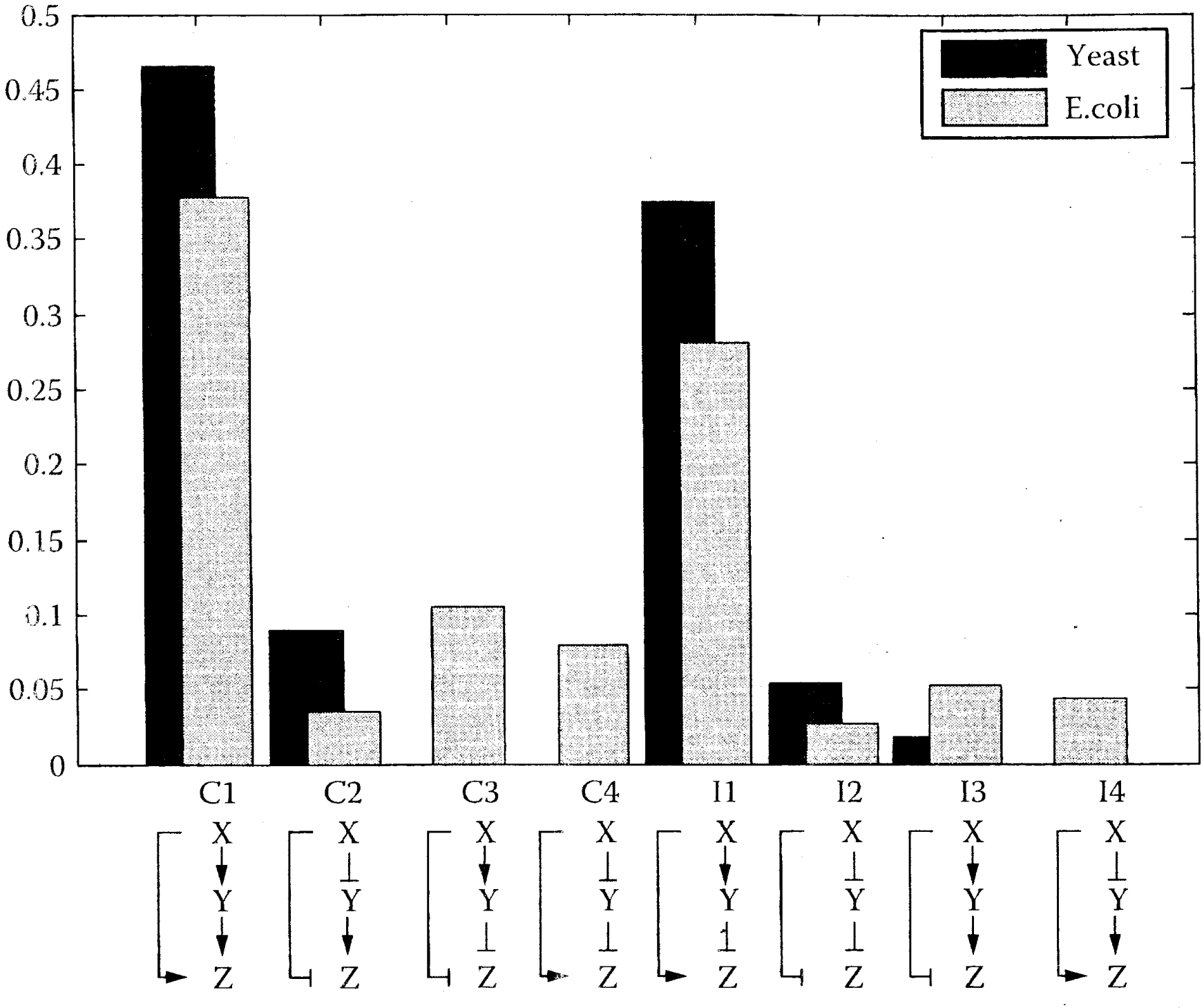
- Made of two cascaded transcription factors (X & Y) that jointly regulate a gene (Z)
- Each of the three interactions of the FFL can be either activation or repression, there are eight types of FFLs
- Motif essentially acts as a low-pass filter, with a time-scale comparable to the delay taken to produce the intermediate protein.

feedforward loop








0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Yeast
E.coli
C1
C2
C3
C4
I1
I2
I3
I4
X
Y
Z

Single-input modules (2/3):

- Single transcription factor controls a set of operons:

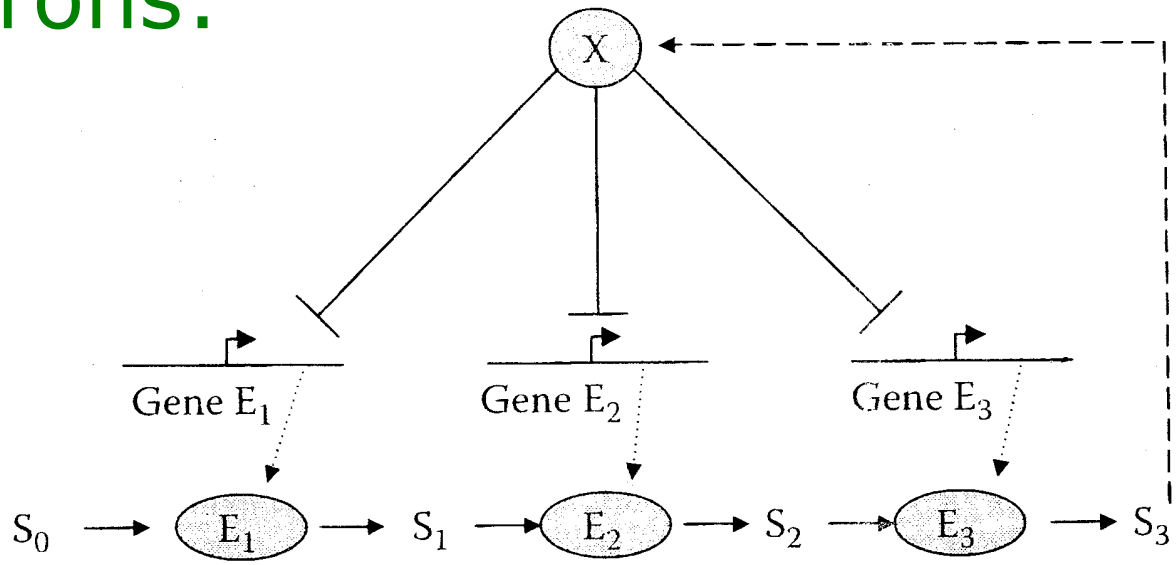
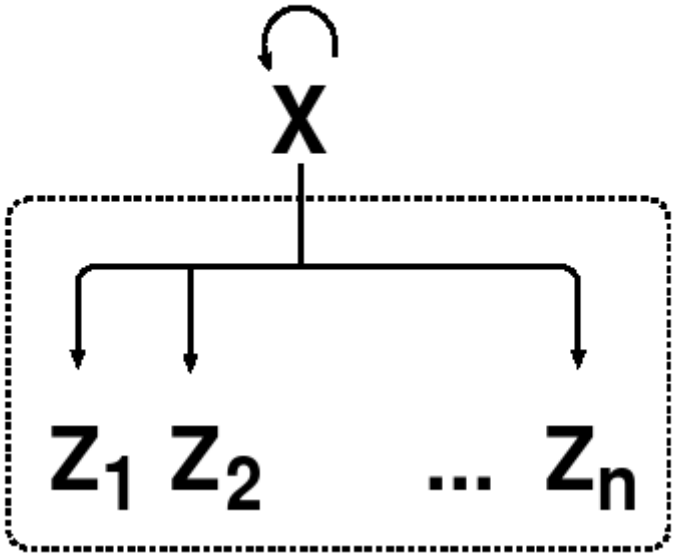
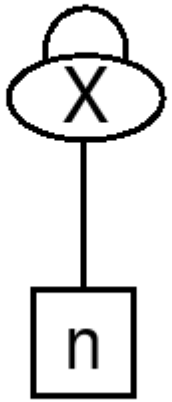
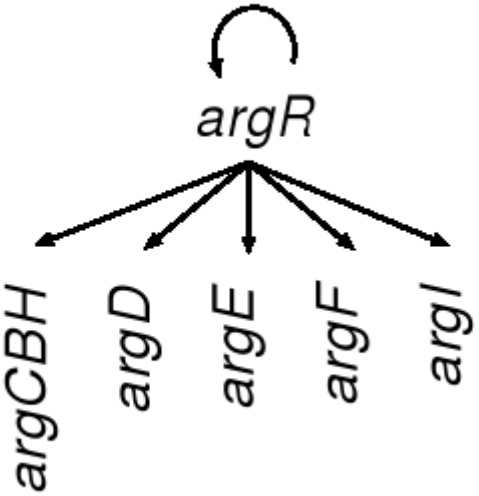


FIGURE 5.2 A single-input module (SIM) regulating a three-step metabolic pathway. The master repressor X represses a group of genes that encode for enzymes E_1, E_2, and E_3 (each on a different operon). These enzymes catalyze the conversion of substrate S_0 to S_1 to S_2, culminating in the product S_3. The product S_3 is the input signal of X: It binds to X and increases the probability that X is in its active state, X*, in which it binds the promoters to repress the production of enzymes. This closes a negative feedback loop, where high levels of S_3 lead to a reduction in its rate of production.

single input module (SIM)



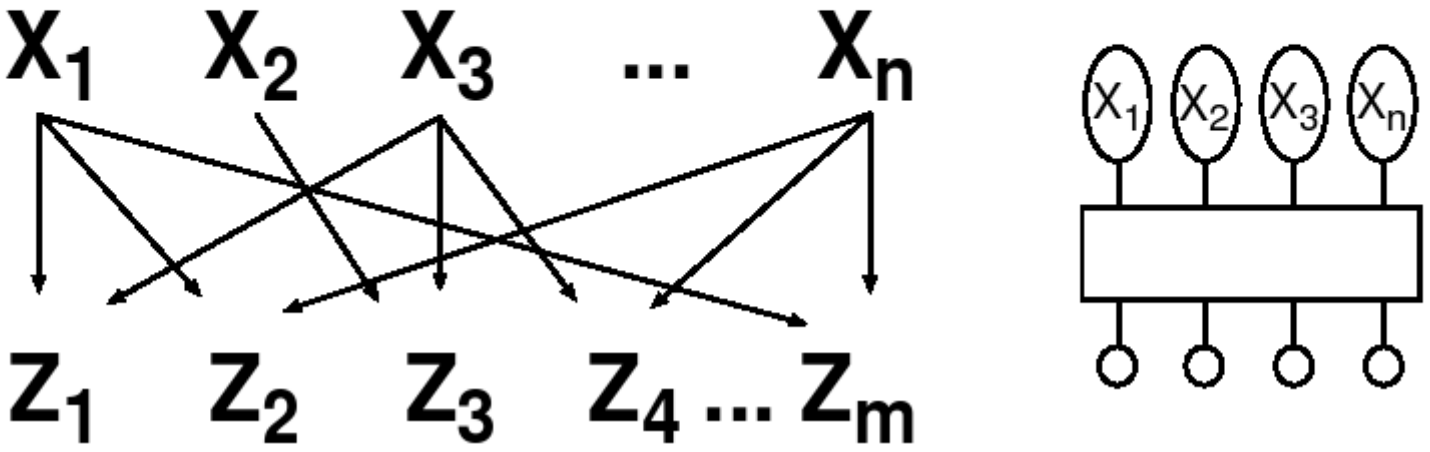


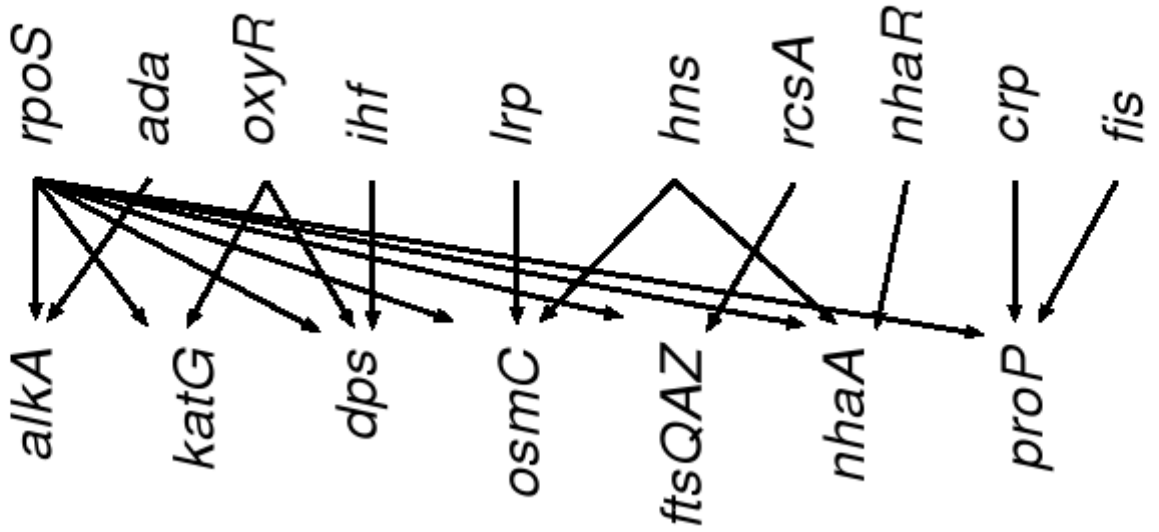


Dense overlapping regulons (3/3):

- Layer of overlapping interactions between operons and a group of input transcription factors
  => Each transcription factor partially controls many operons
- Much more dense (overlapping) than corresponding structures in randomized networks
- Found by clustering TFs that operate on same operons

dense overlapping regulons (DOR)





Network dynamics:

- Network motifs may have specific functions in the information processing performed by the network
- One clue to their possible function is provided by common themes of the systems in which they appear
- Insight may be gained by mathematical analysis of their dynamics
- Boolean kinetics were used, but only for demonstration purposes:

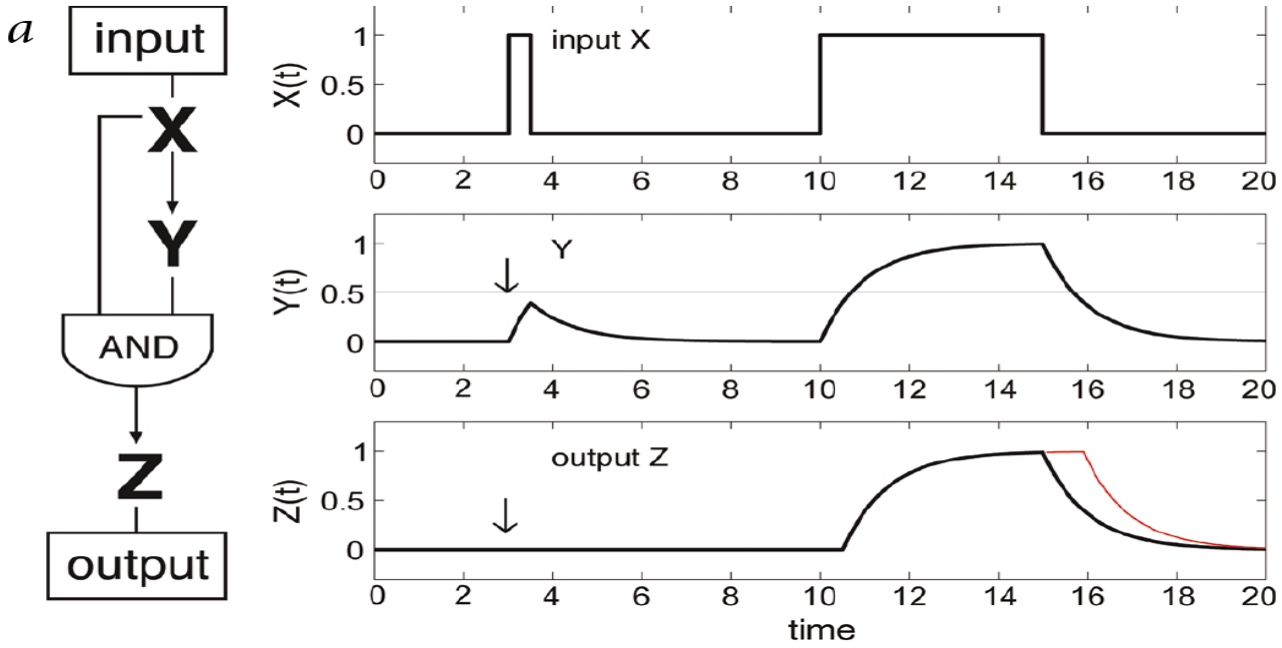
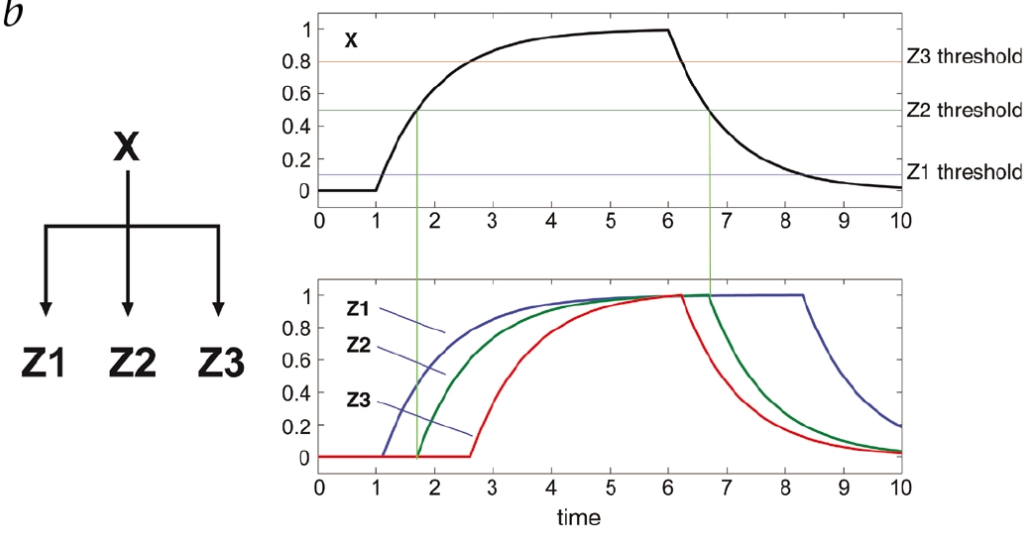


Fig. 2 Dynamic features of the coherent feedforward loop and SIM motifs. ***a***, Consider a coherent feedforward loop circuit with an 'AND-gate'–like control of the output operon Z. This circuit can reject rapid variations in the activity of the input X, and respond only to persistent activation profiles. This is because Y needs to integrate the input X over time to pass the activation threshold for Z (thin line). A similar rejection of rapid fluctuations can be achieved by a cascade, X→Y→Z; however, the cascade has a slower shut-down than the feedforward loop (thin red line in the Z dynamics panel). ***b***, Dynamics of the SIM motif. This motif can show a temporal program of expression according to a hierarchy of activation thresholds of the genes. When the activity of X, the master activator, rises and falls with time, the genes with the lowest threshold are activated earliest and deactivated latest. Time is in units of protein lifetimes, or of cell cycles in the case of long-lived proteins.



(Their) Conclusions:

- Motifs describe transcriptional network in compact and modular form
- Only limited set of transcription interactions used ("slow part of regulatiory network")
- Faster (protein-protein) regulation exists
- The concept of network motifs still holds and (with dynamics) can increase understanding of transcriptional networks (and others)

“Easily interpretable view”

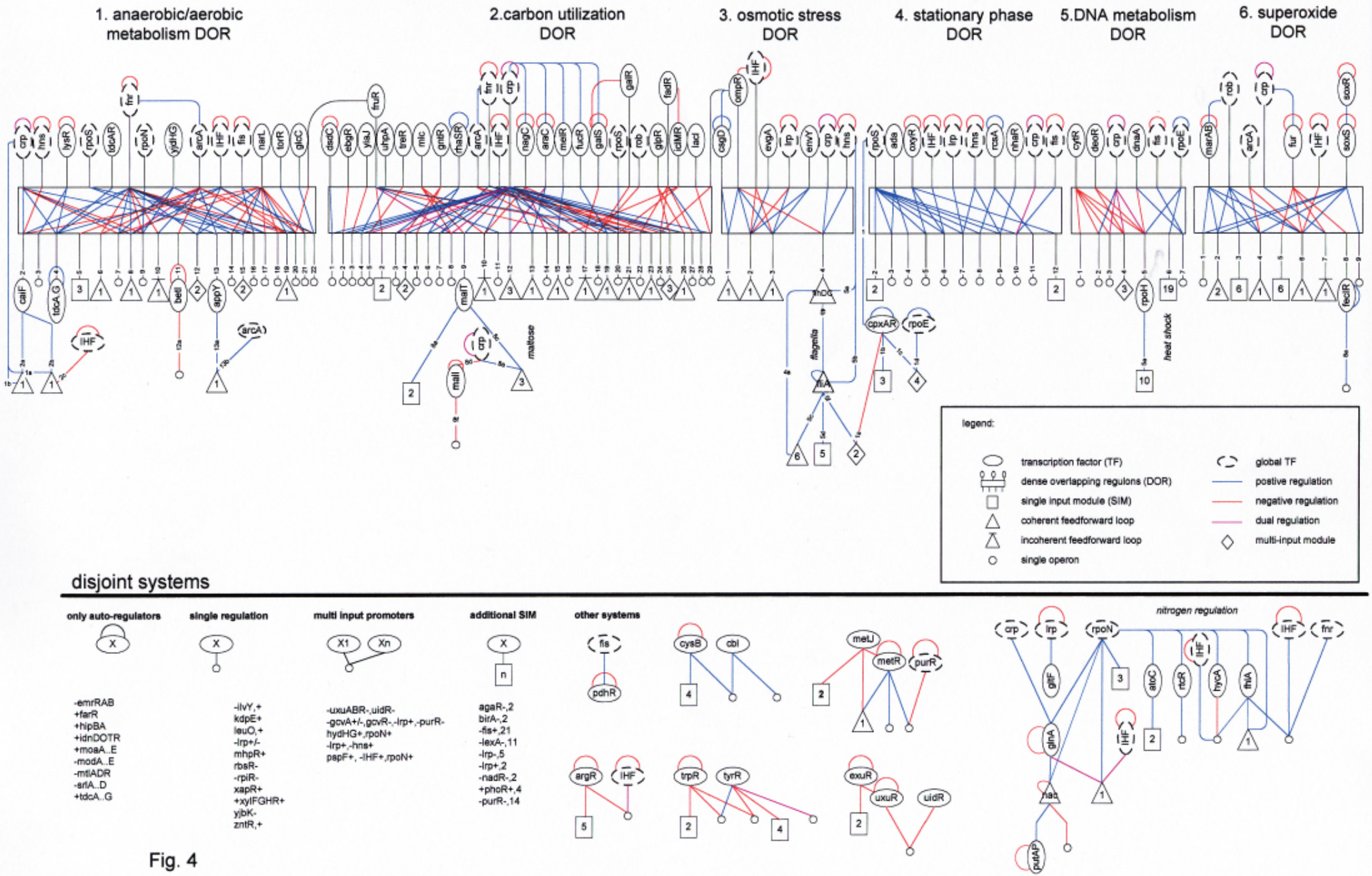


Fig. 4

Critique / Problems (1/3):

- The main problem seems to be the null-hypothesis (ensemble of random networks as a baseline for comparisons)
- For example, simple connectivity does not take into account spatial relationships (Artzy-Randrup et.al:”C. elegans & neurons”)

=> random flipping of edges does not generate “proper” random networks

Critique / Problems (2/3):

- Authors claim that the network motifs are so important that they are preserved during the evolution (even:”one of the governing principles of evolution”)
- Others (Meshi et.al, Mazurie et.al.) have found that although network motifs seem to exists, they are not:
  - overly conserved during the evolution
  - do not seem to correlate to functional enrichments (functions are conserved)

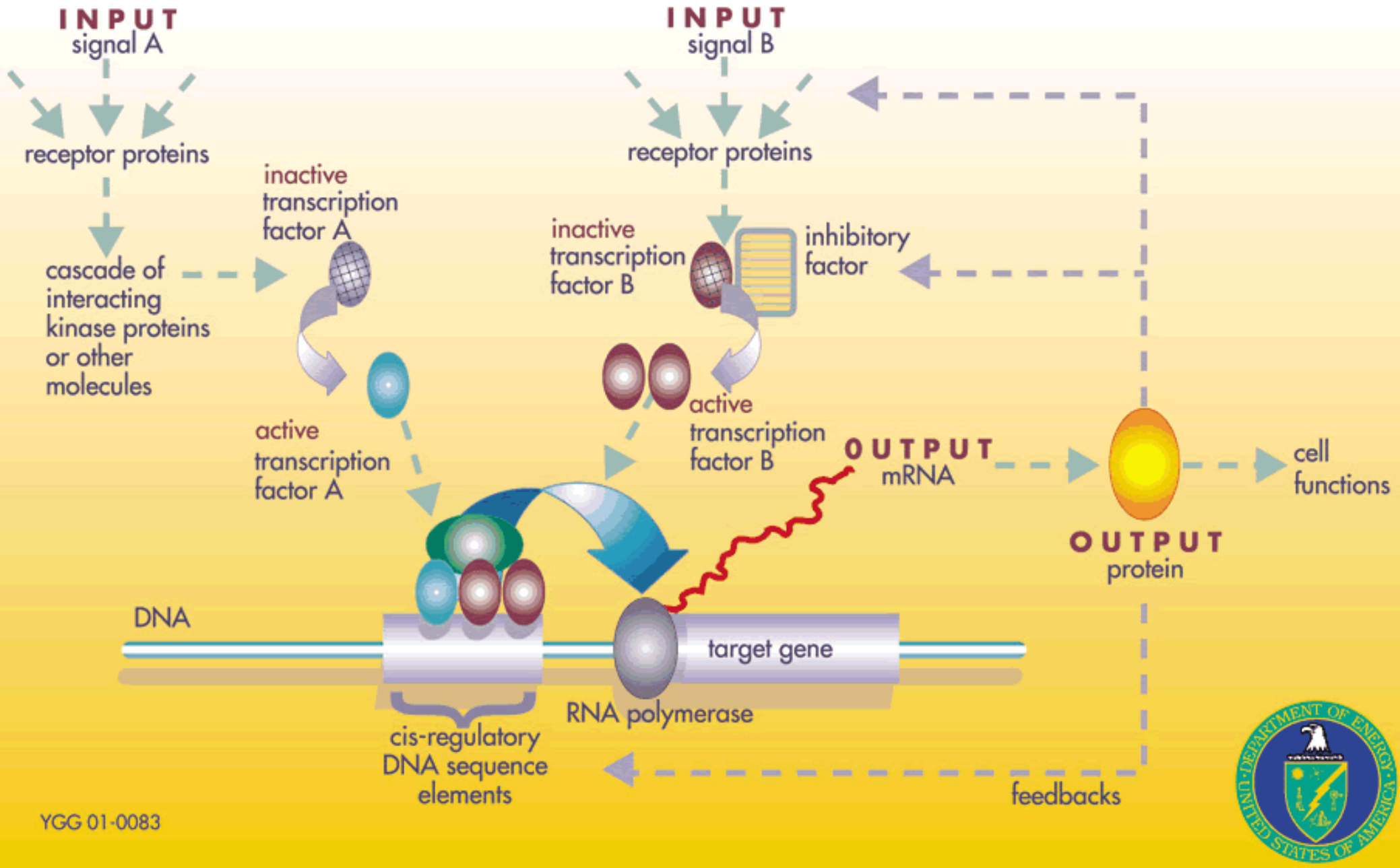
Critique / Problems (3/3):

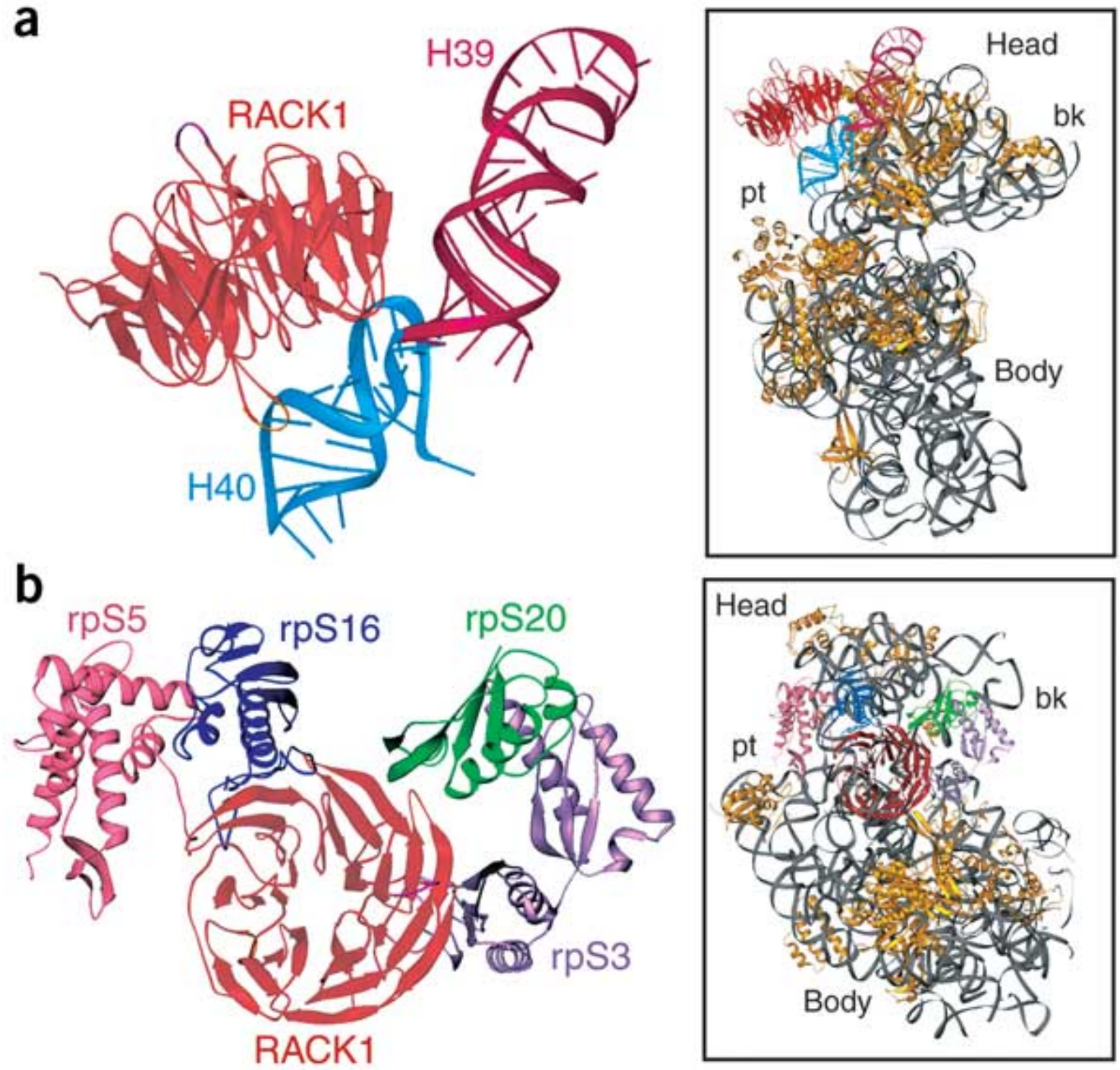
- Like small world hypotheisis – applies to anything and everything...?
- Forgets all post-translational modifications of proteins
- Evolution is not seeking optimum (vs. engineering)
- Essentially none of the motifs exists in isolation (time / space) - topological findings only

Conclusions:

- Network motifs may provide a systems level view into cell's life-cycle:
  - (topological) building blocks
  - network dynamics
- Problems with the definition of null-hypothesis (random networks)
- Don't necessarily explain anything new
- Oversimplification vs. predictive power? (Could work with large enough motifs?)

A GENE REGULATORY NETWORK




a
H39
RACK1
H40
Head
bk
pt
Body
b
rpS5
rpS16
rpS20
rpS3
RACK1
Head
bk
pt
Body

The End / Thank You!