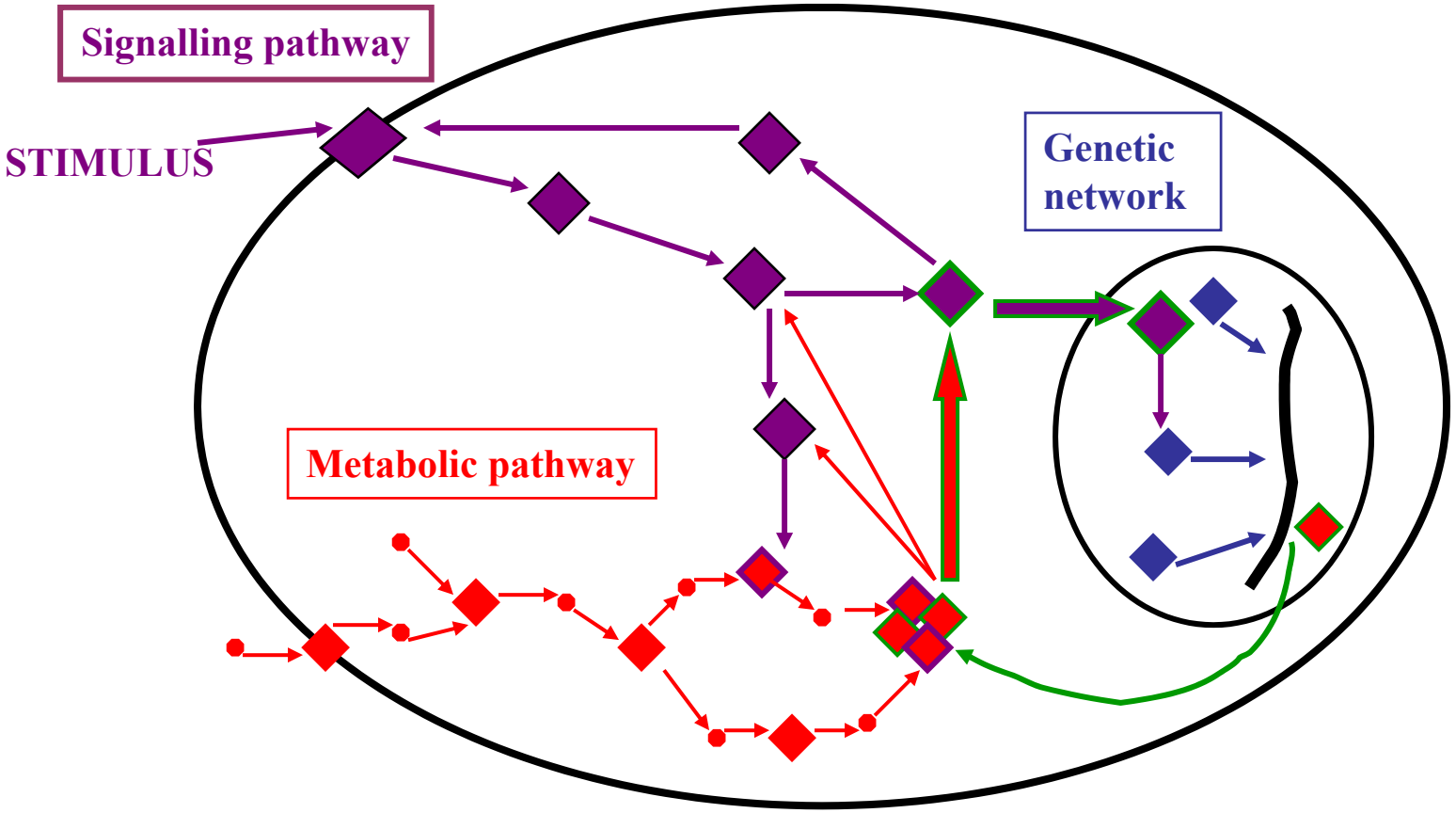
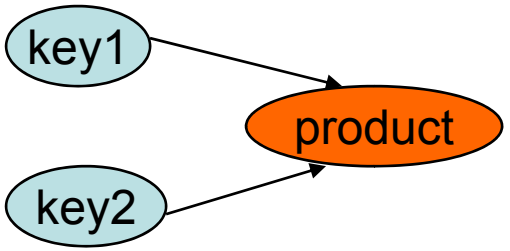
Biological networks

Pathways are inter-linked



Gene regulatory networks

- Transcription Factors: special proteins that function as “keys” to the “switches” that determine whether a protein is to be produced
- Gene regulatory networks try to show this “key-product” relationship and understand the regulatory mechanisms that govern the cell.



Other biological networks?

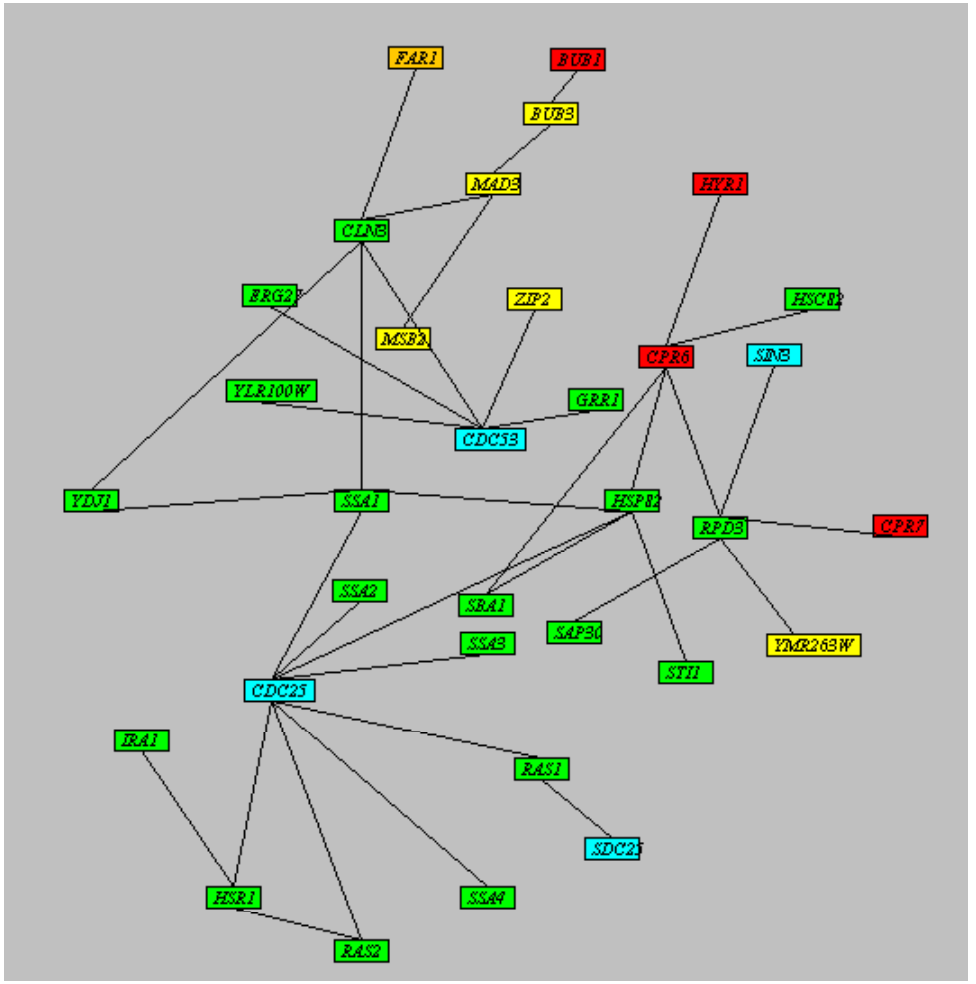
- Apart from regulation there are other events in a cell that require interaction of biological molecules
- Other types of molecular interactions that can be observed in a cell
  - enzyme – ligand
    - **enzyme:** a protein that catalyzes, or speeds up, a chemical reaction
    - **ligand:** extracellular substance that binds to receptors
    - metabolic pathways
  - protein – protein
    - cell signaling pathways
    - proteins interact physically and form large complexes for cell processes

Interactions → Pathways → Network

- A collection of interactions defines a network
- Pathways are subsets of networks
  - All pathways are networks of interactions, however not all networks are pathways!
  - Difference in the level of annotation or understanding
- We can define a pathway as a biological network that relates to a **known** physiological process or complete function

The “interactome”

- The complete wiring of a proteome.
- Each vertex represents a protein.
- Each edge represents an “interaction” between two proteins.



An edge between two proteins if...

- The proteins interact physically and form large complexes
- The proteins are enzymes that catalyze two successive chemical reactions in a pathway
- One of the proteins regulates the expression of the other

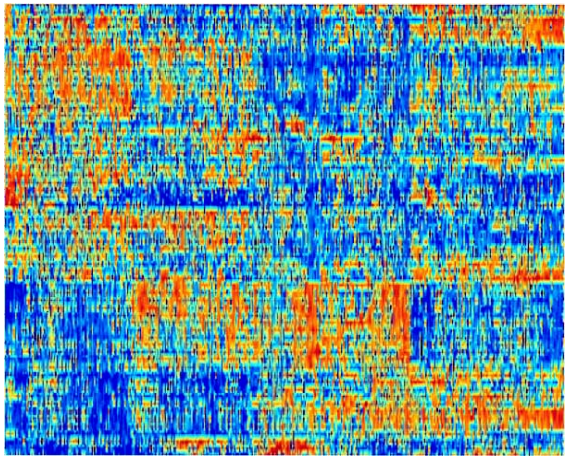
Gene Regulatory Networks

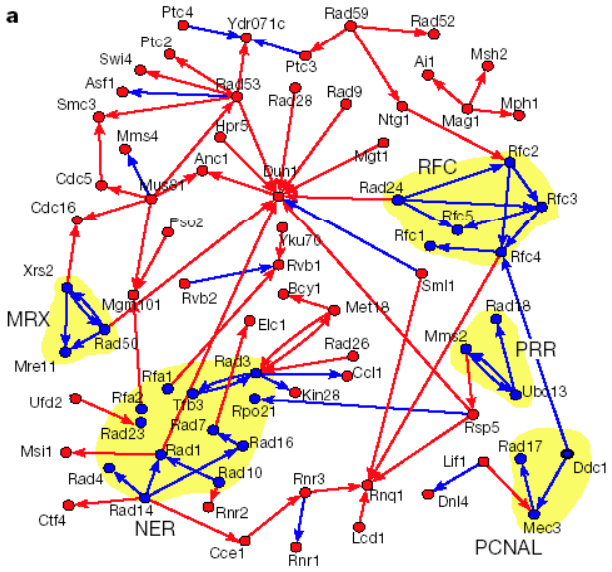
slides adapted from

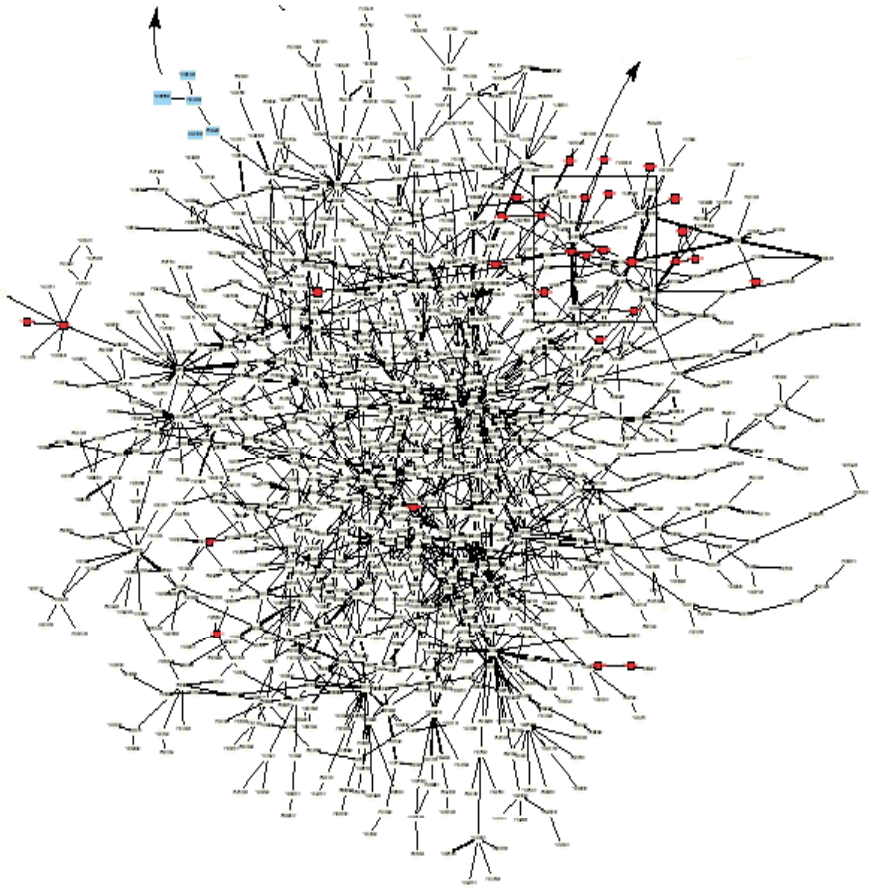
Shalev Itzkovitz's talk

given at IPAM UCLA on July 2005

Protein networks - optimized molecular computers





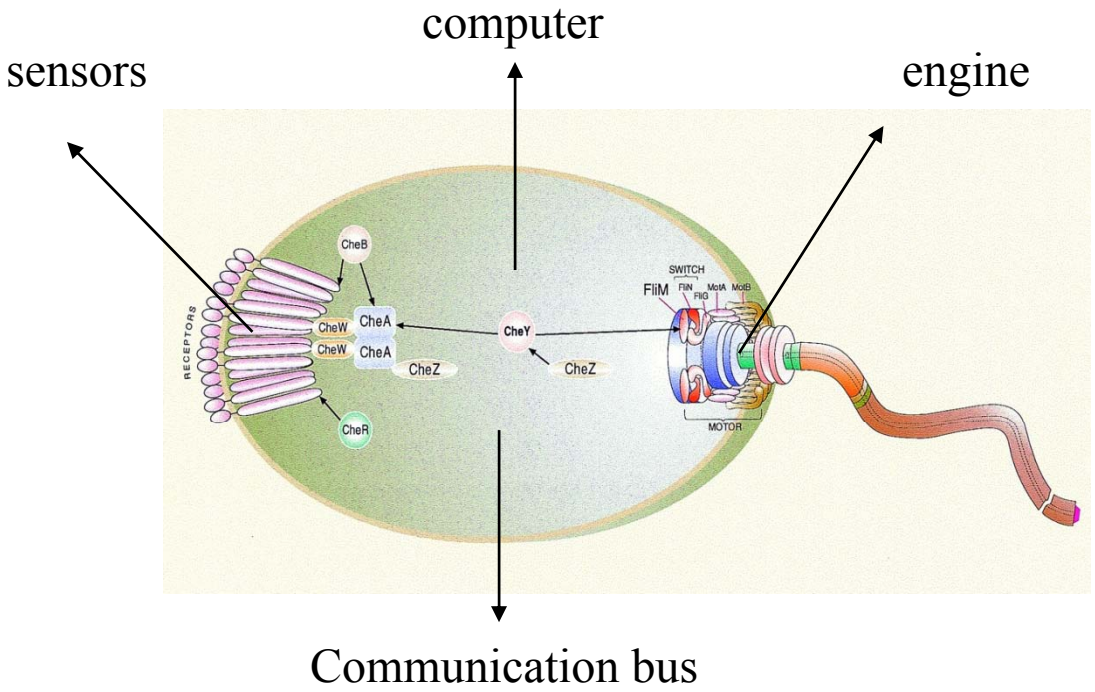


E. coli – a model organism

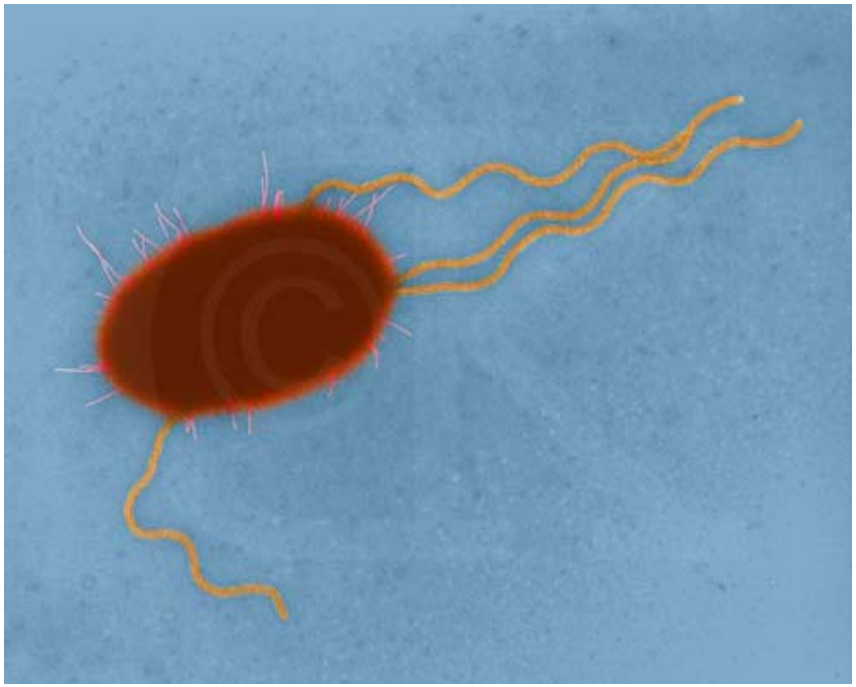
Single cell, 1 micron length

Contains only ~1000 protein types at any given moment

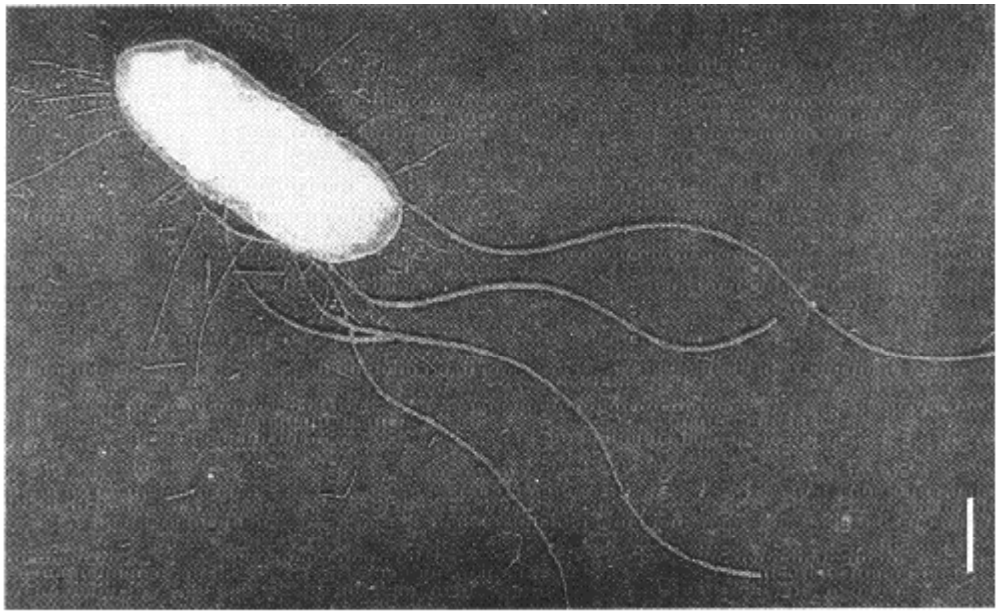
still : **Amazing technology**

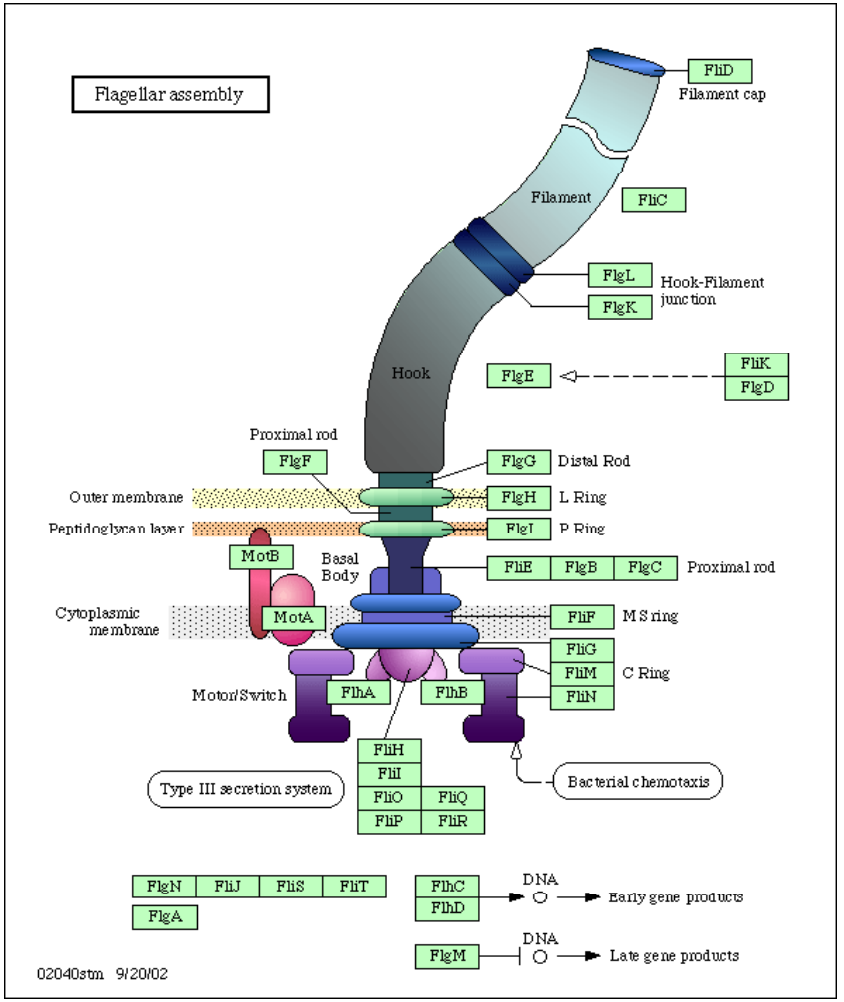


Can move toward food and away from toxins



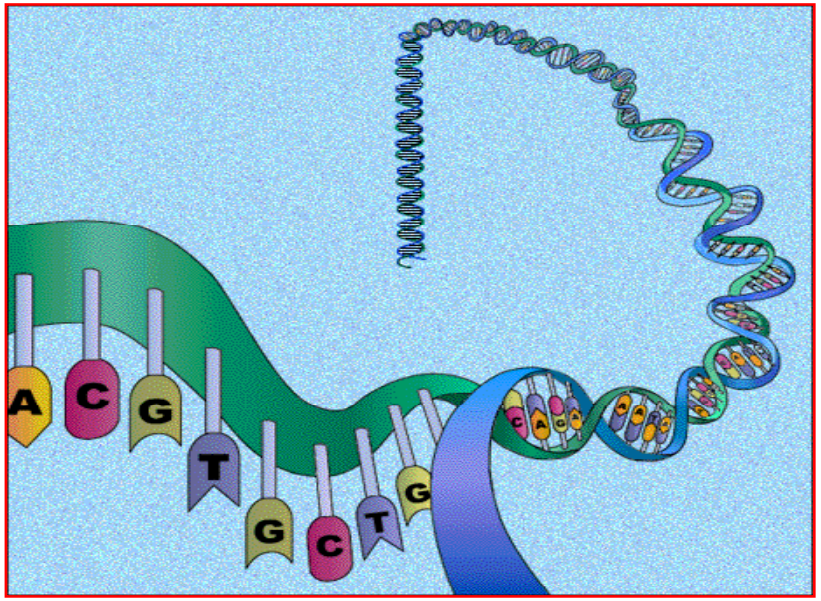
Flagella assembly

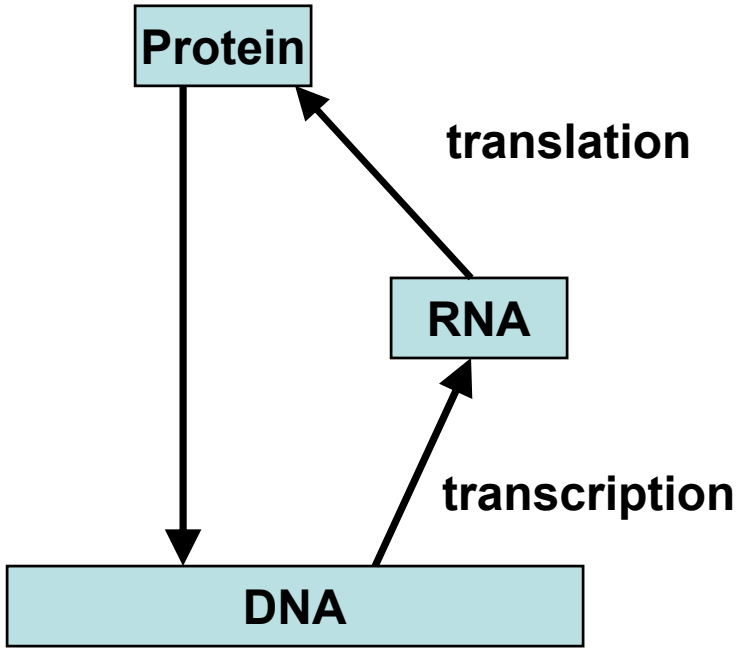




- Composed of 12 types of proteins
- Assembled only when there is an environmental need for motility
- Built in an efficient and precise temporal order

Proteins are encoded by DNA





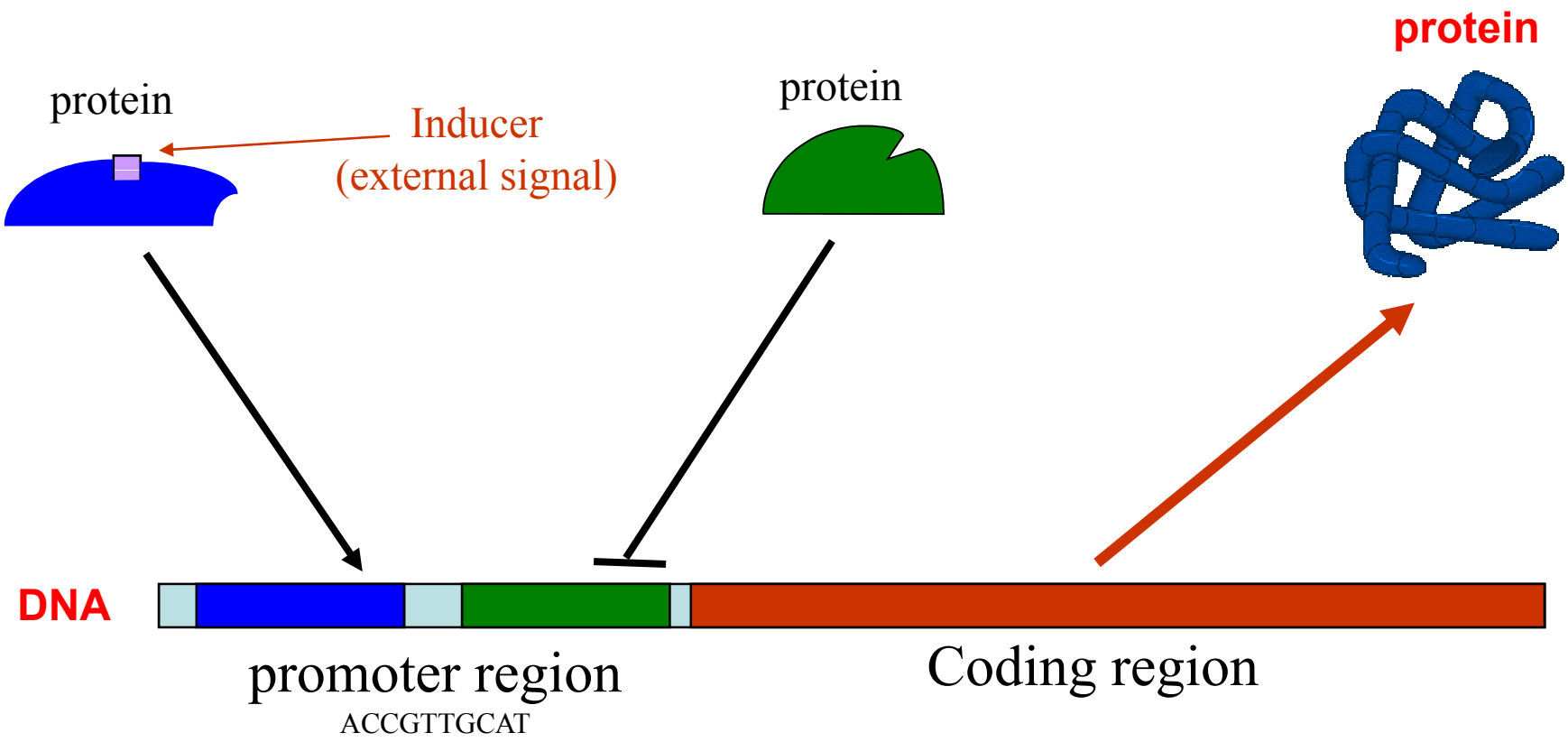
DNA – same inside every cell, the instruction manual, 4-letter chemical alphabet – A,G,T,C

E. Coli – 1000 protein types at any given moment

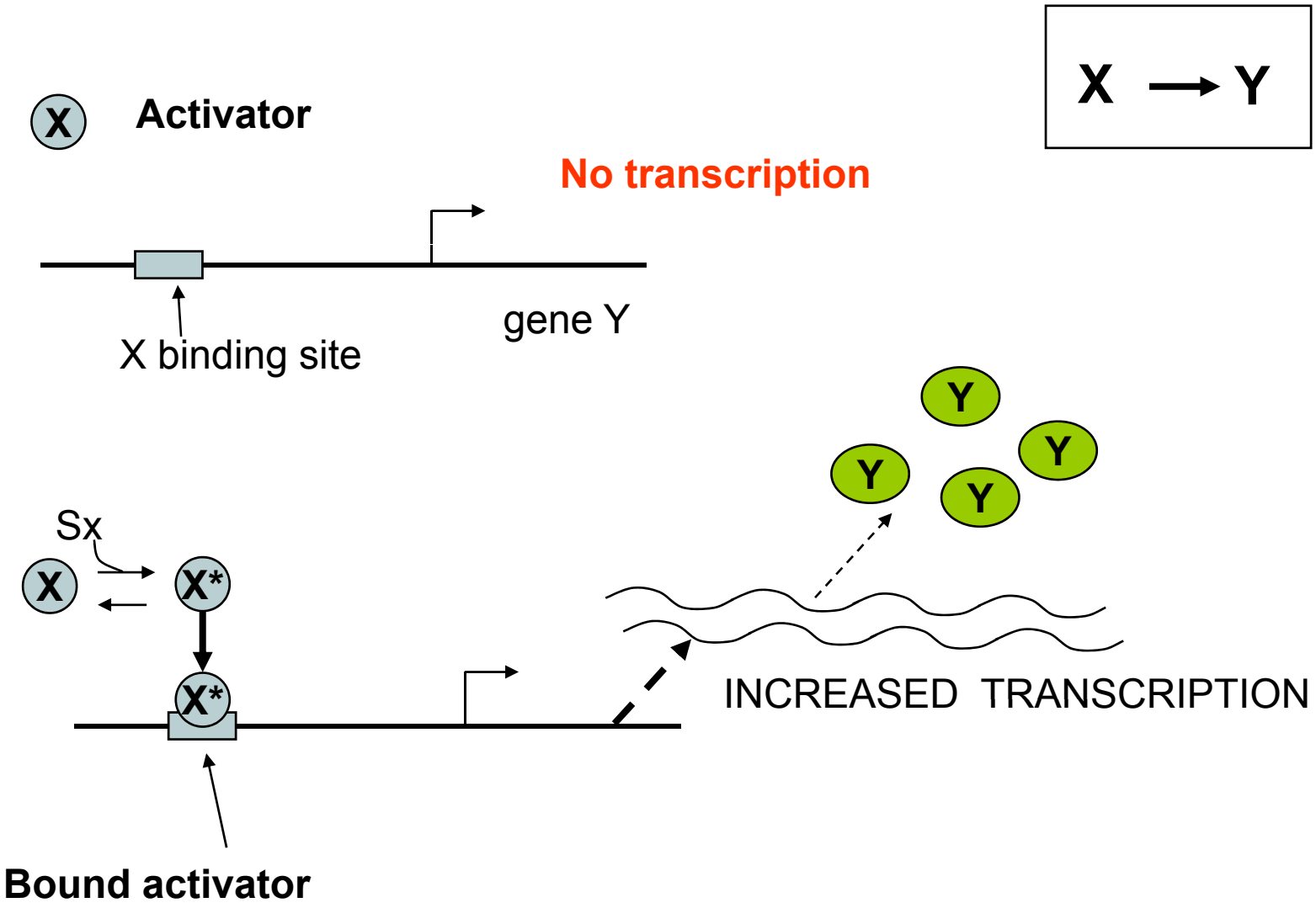
>4000 genes (or possible protein types) – need regulatory mechanism to select the active set

Gene Regulation

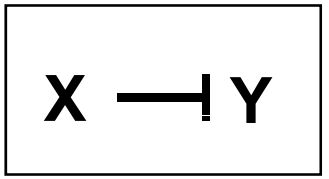
- Proteins are encoded by the DNA of the organism.
- **Proteins regulate expression of other proteins by interacting with the DNA**



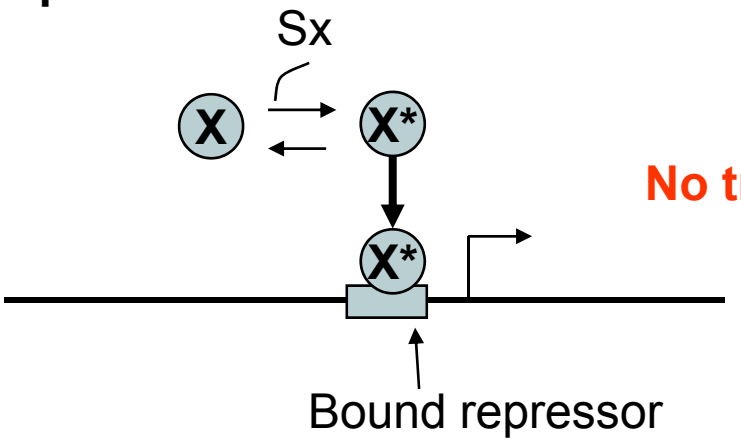
Activators increase gene production



Repressors decrease gene production

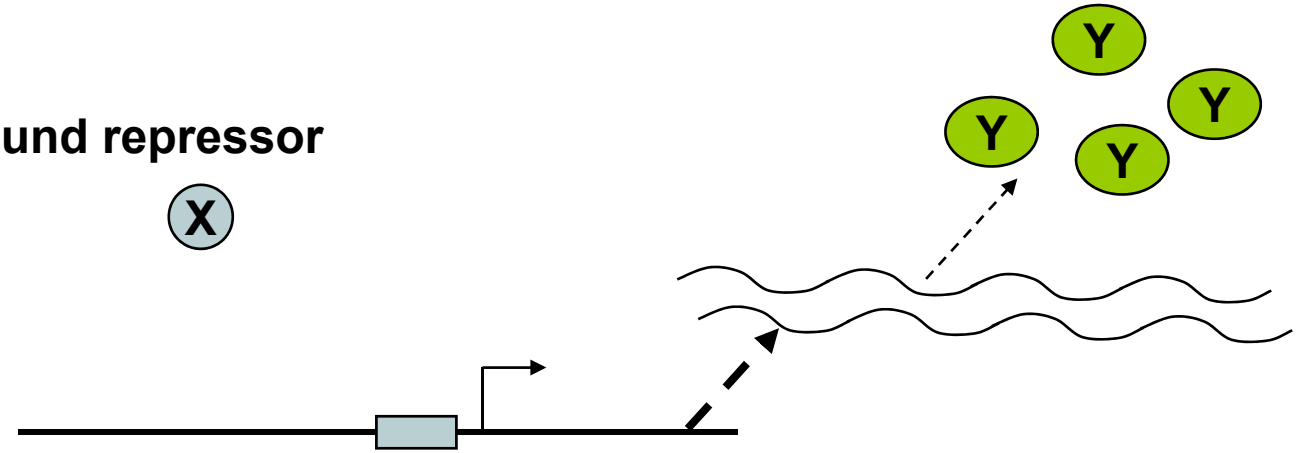


Bound repressor

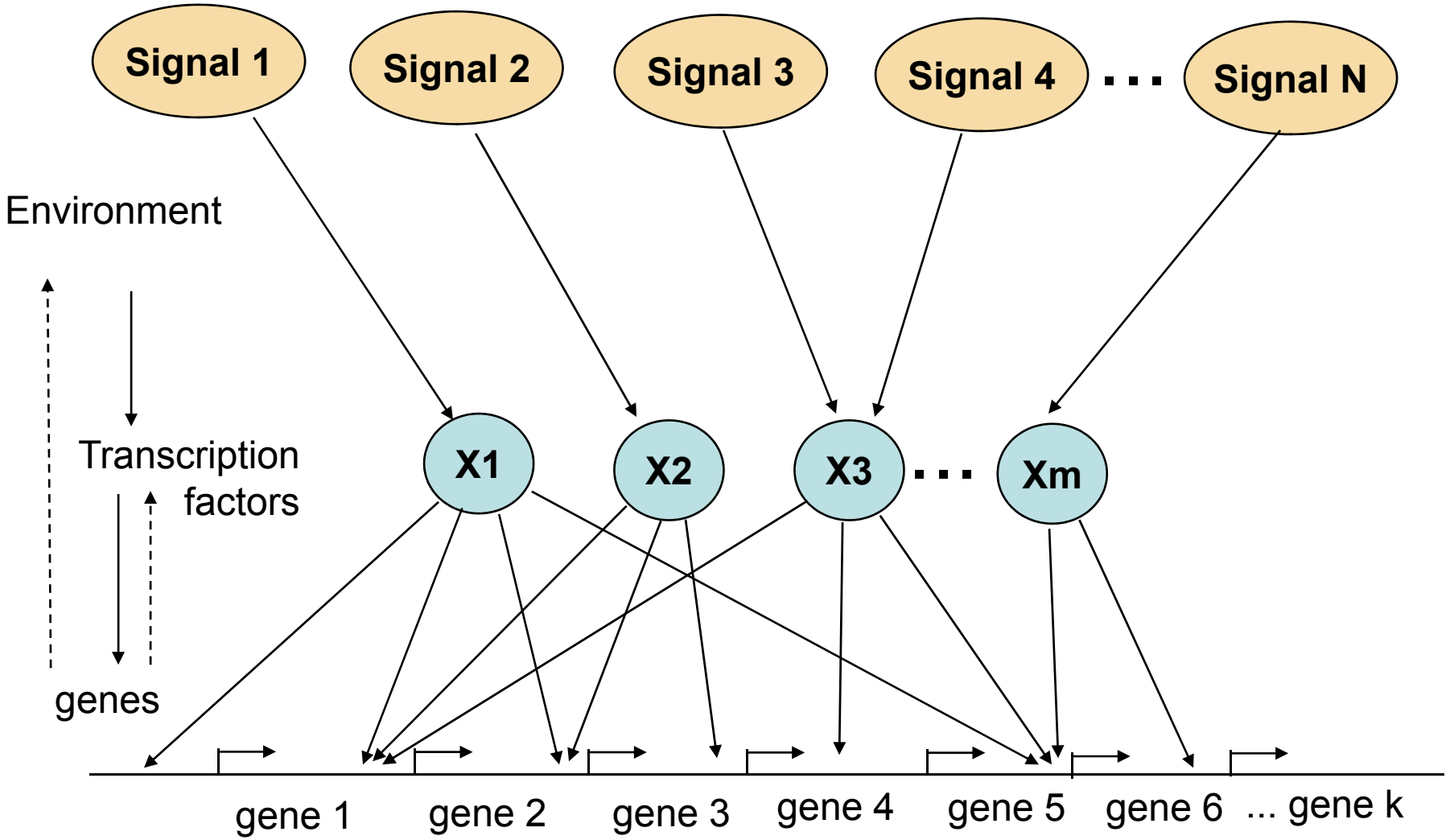


No transcription

Unbound repressor

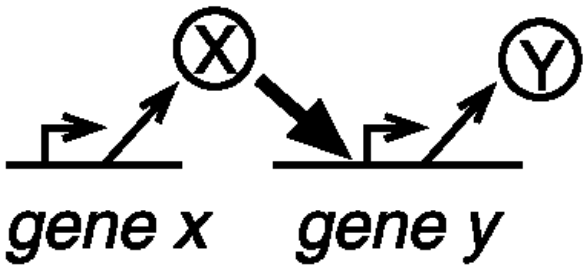


An environmental sensing mechanism

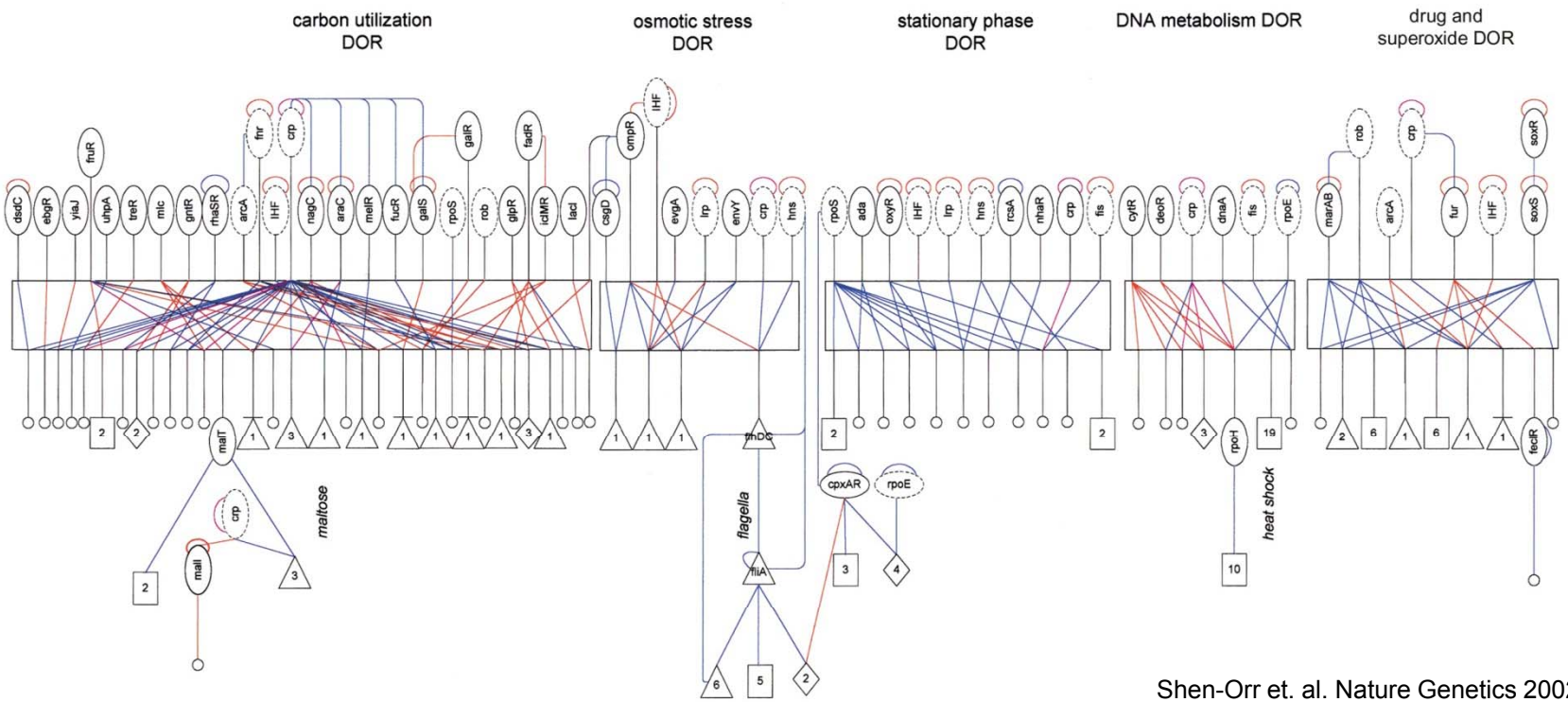


Gene Regulatory Networks

- Nodes are proteins (or the genes that encode them)



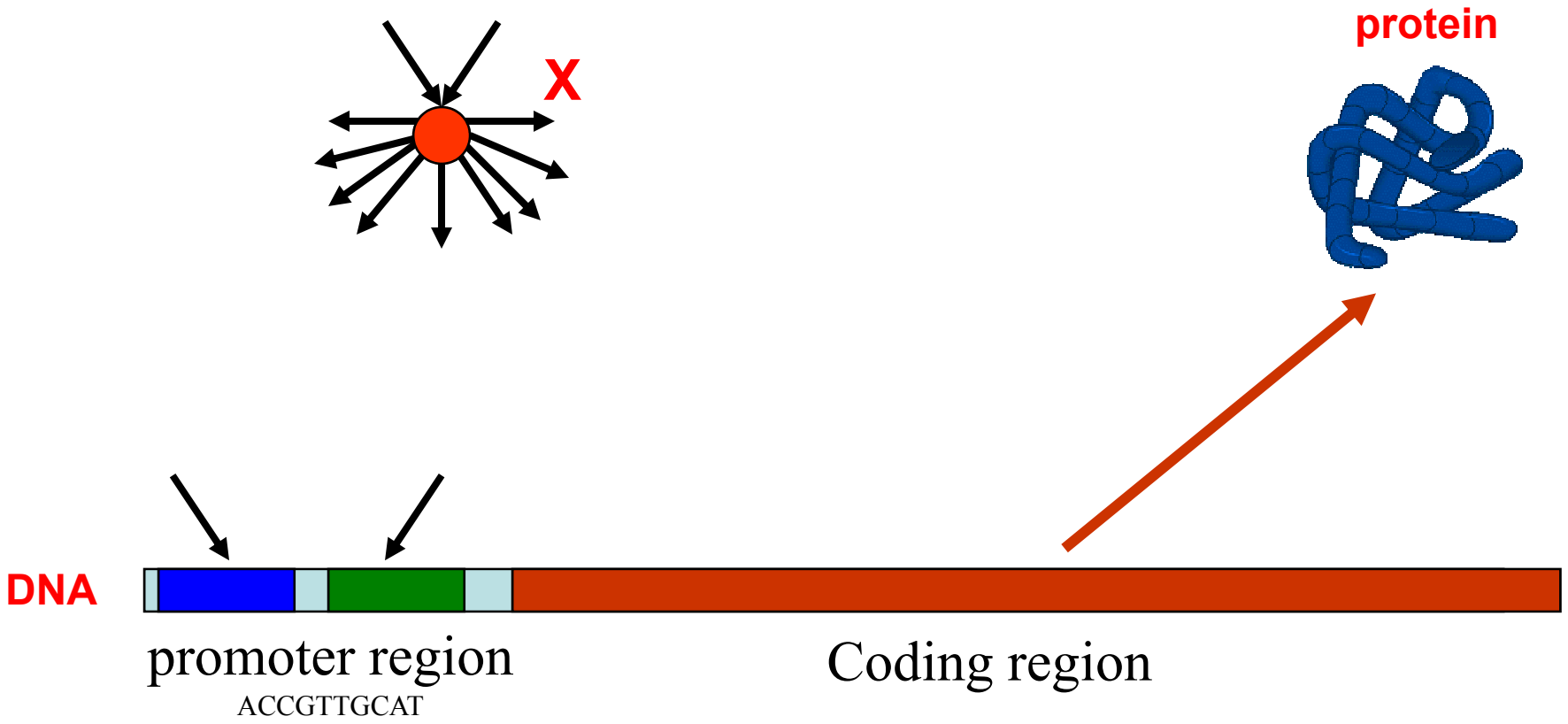
The gene regulatory network of E. coli



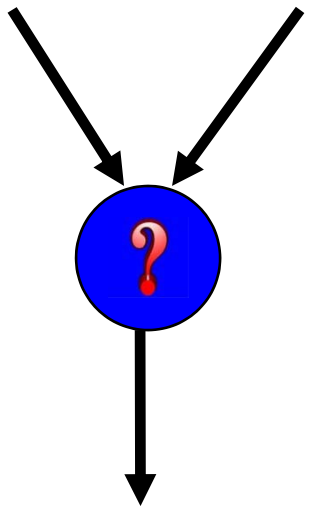
Shen-Orr et. al. Nature Genetics 2002

- **shallow network**, few long cascades.
- modular
- compact in-degree (promoter size limitation)

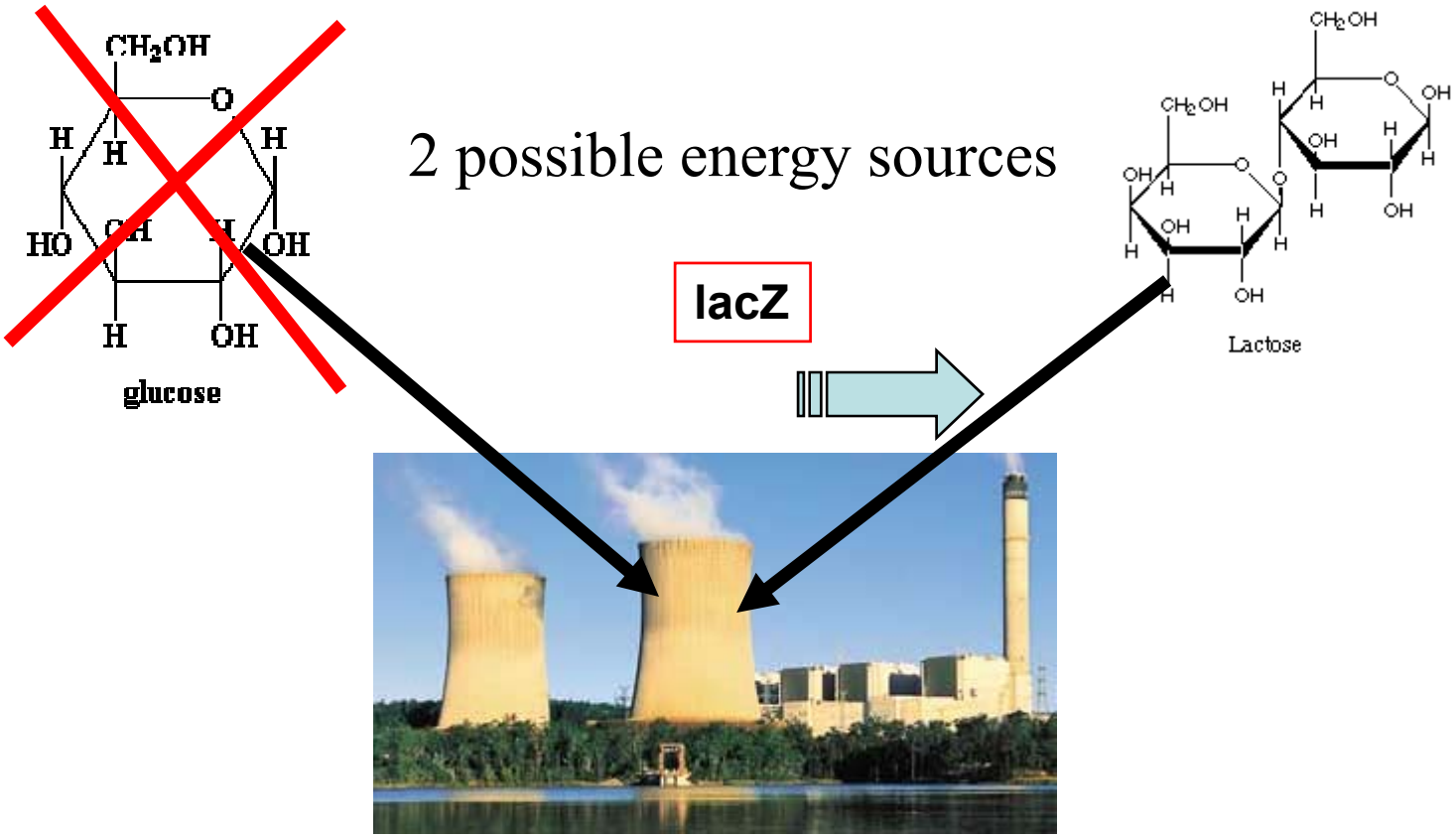
Asymmetric degree distribution due to Promoter size limitation



What logical function do the nodes represent?



Example – Energy source utilization



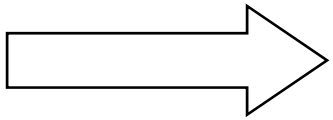
The E. coli prefers glucose

lacZ is a protein needed to break down lactose into carbon

How will the E. coli decide when to create this protein?

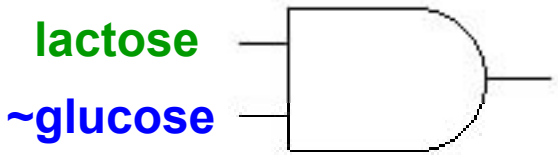
Proteins have a cost

- E. Coli creates ~10^6 proteins during its life time
- ~1000 copies on average for each protein type



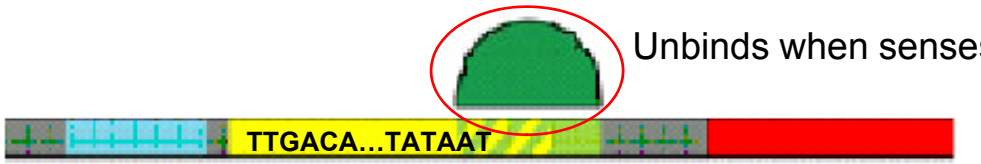
E. Coli will grow 1/1000 slower,
Enough for evolutionary pressure

AND gate encoded by proteins and DNA



lacZ gene is controlled by 2 “sensory” proteins :

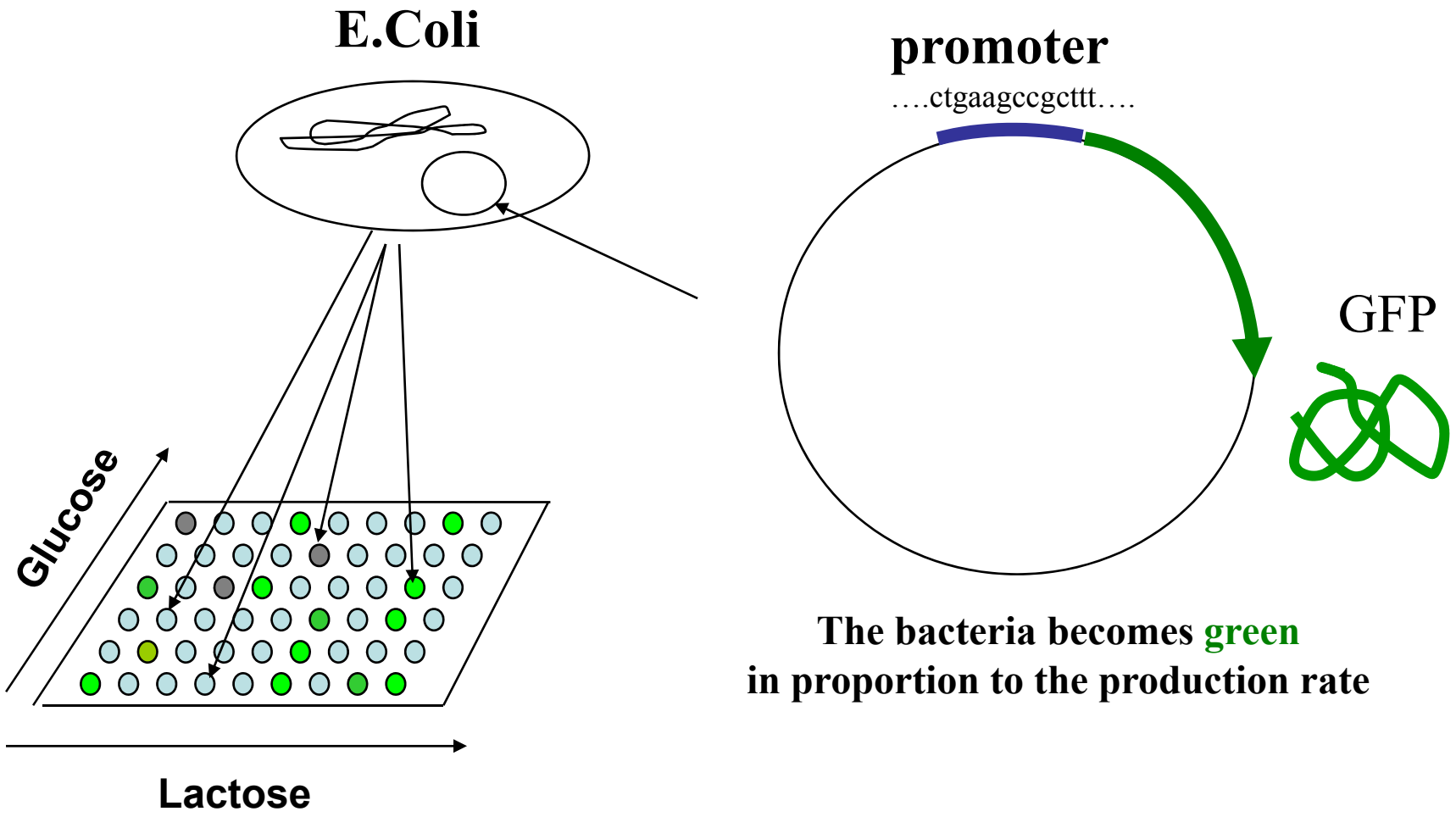
lactose sensor



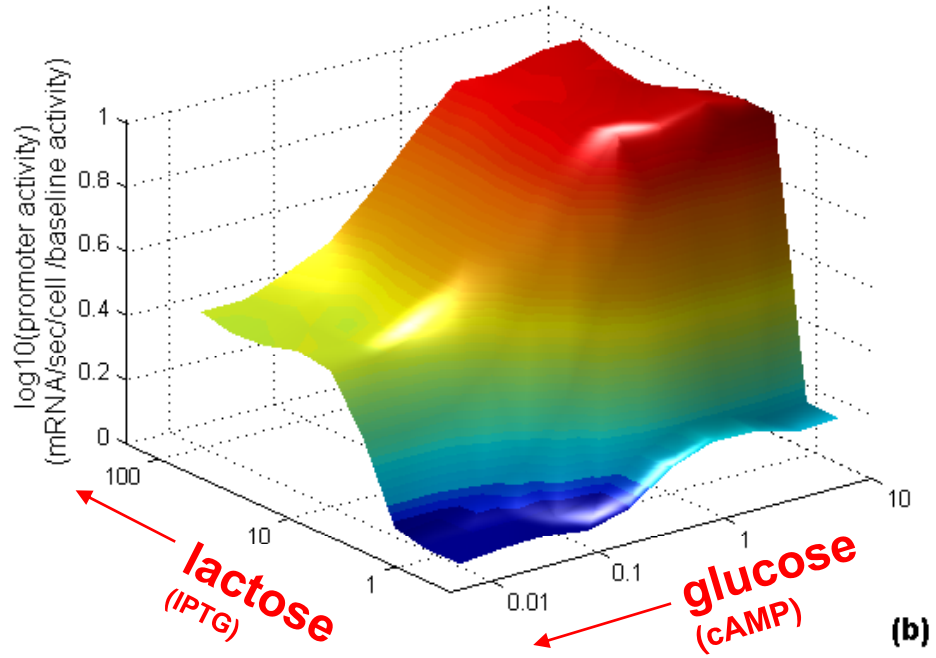
glucose absence sensor

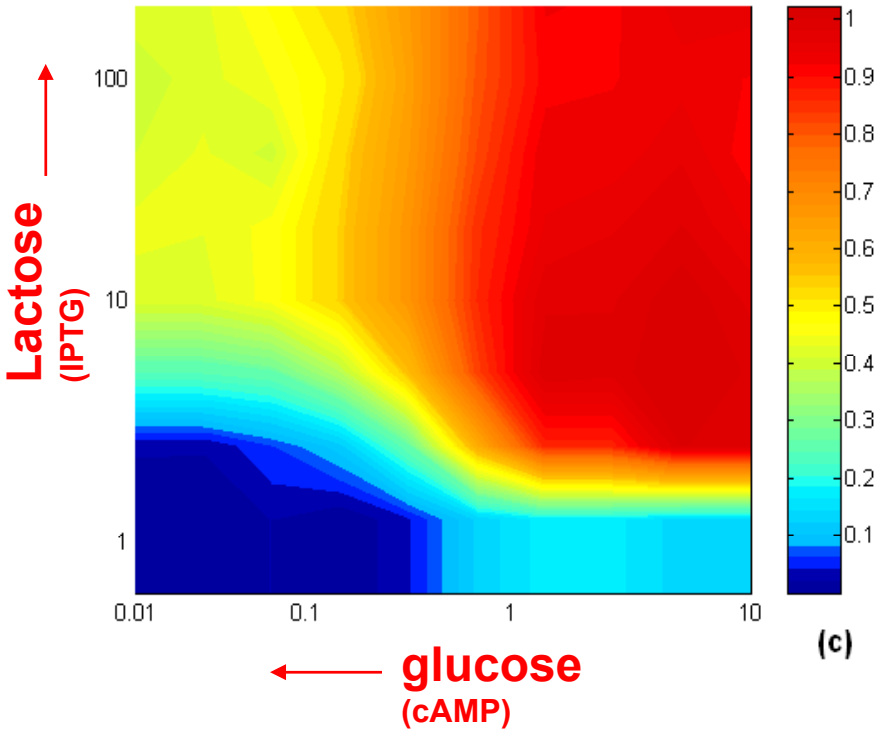
Jacob & Monod, J. Mol. Biol. 1961

Experimental measurement of input function



The input function of the lactose operon is more elaborate than a simple AND gate

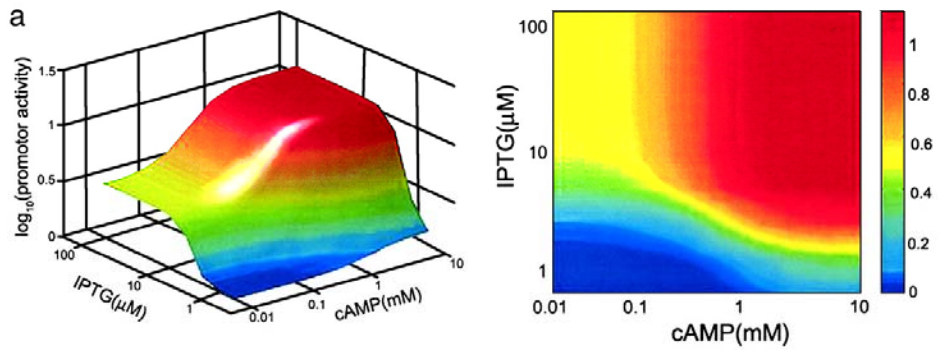




Setty et. al. PNAS 2003

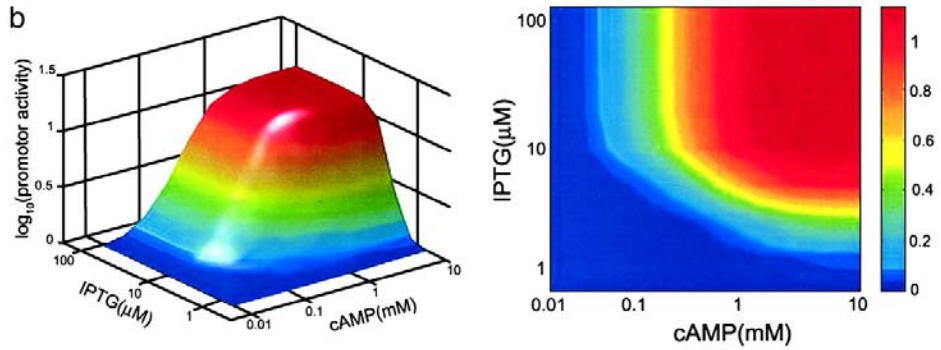
E. Coli can modify the input function by small changes in the promoter DNA

…AAGGCCT…



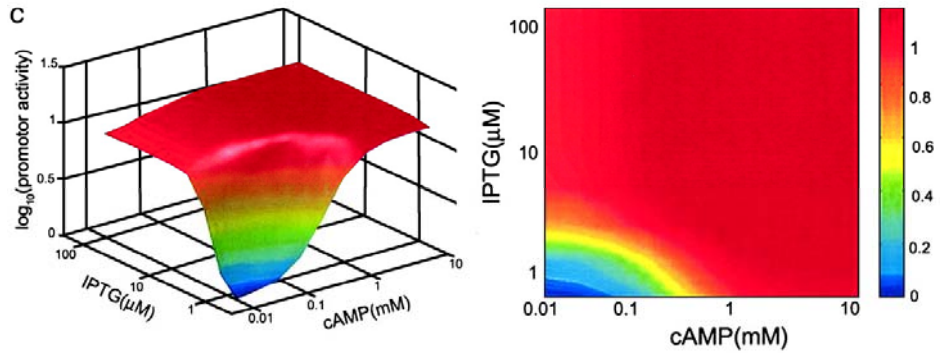
LacZ gate

…AAG**T**CCT…



AND gate

…AAG**T**C**T**T…

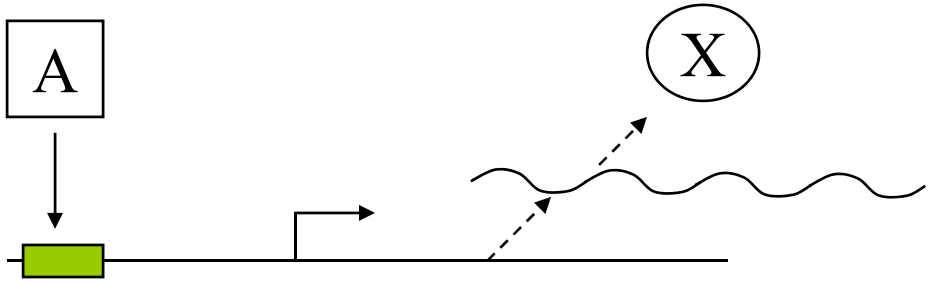


OR gate

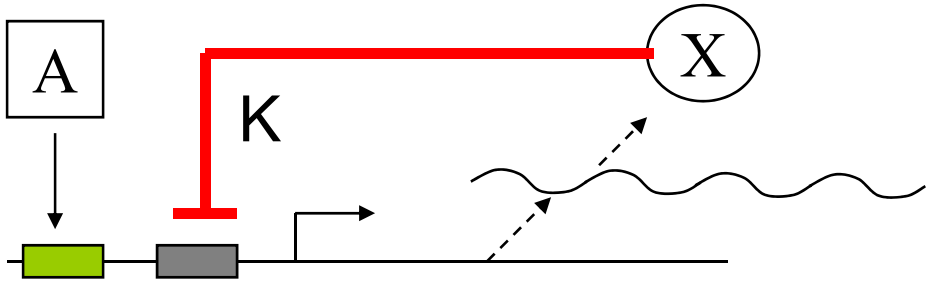
Input function is optimally tuned to the environment

Negative autoregulation

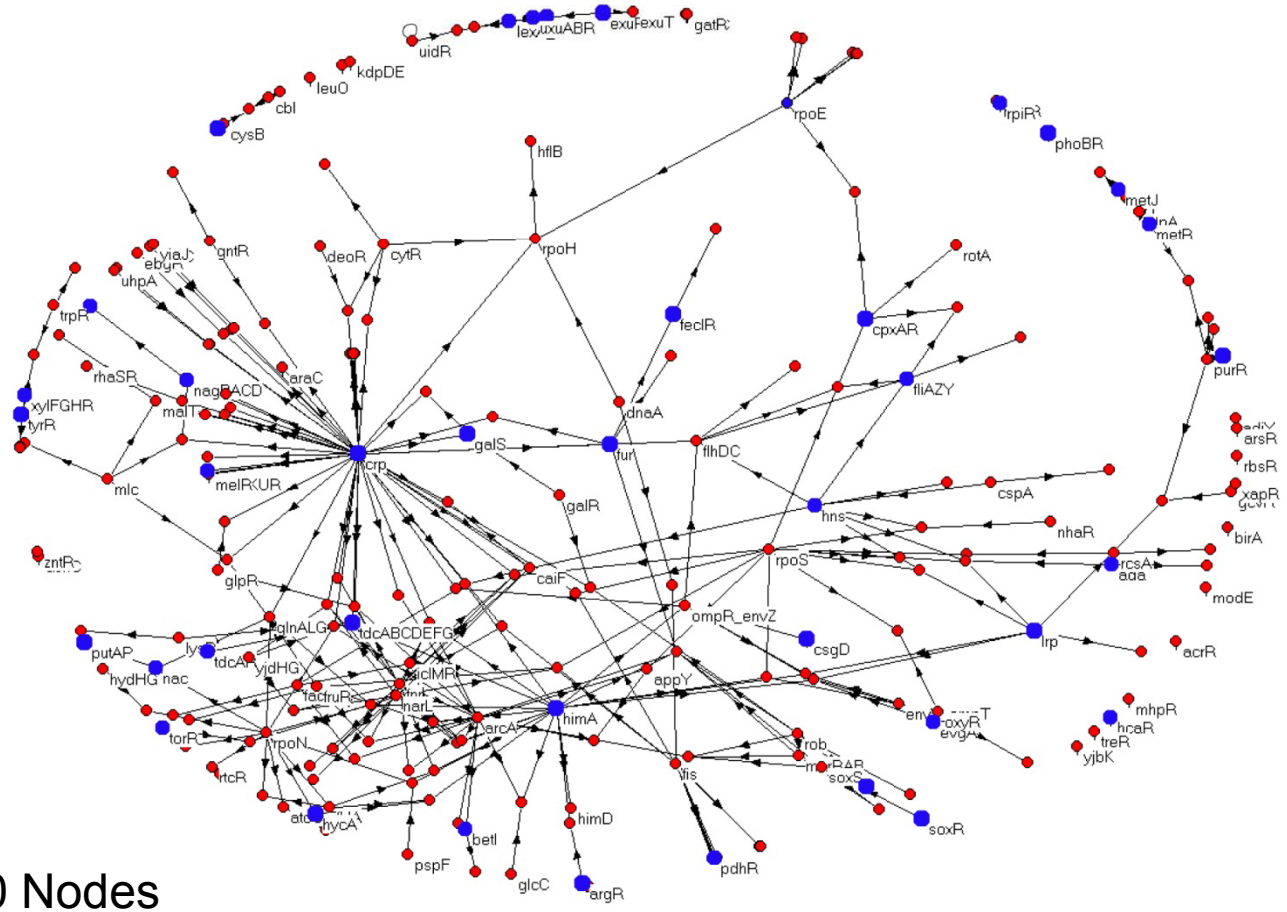
Simple regulation



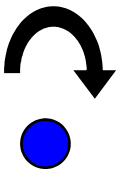
Negative autoregulation



Negative autoregulation is a hugely statistically significant pattern



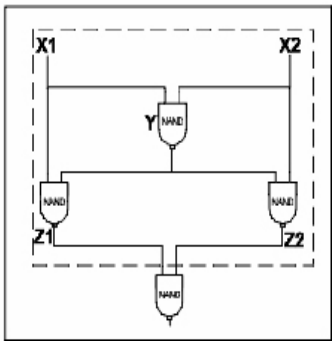
N=420 Nodes
E=520 Edges
Es=40 self-edges

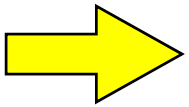


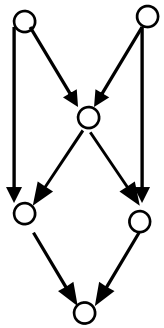
Blue nodes have self-edges

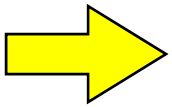
A protein with negative autoregulation is a recurring pattern with a defined function

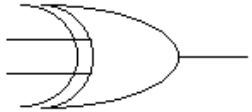
Are there larger recurring patterns which play a defined functional role ?











XOR

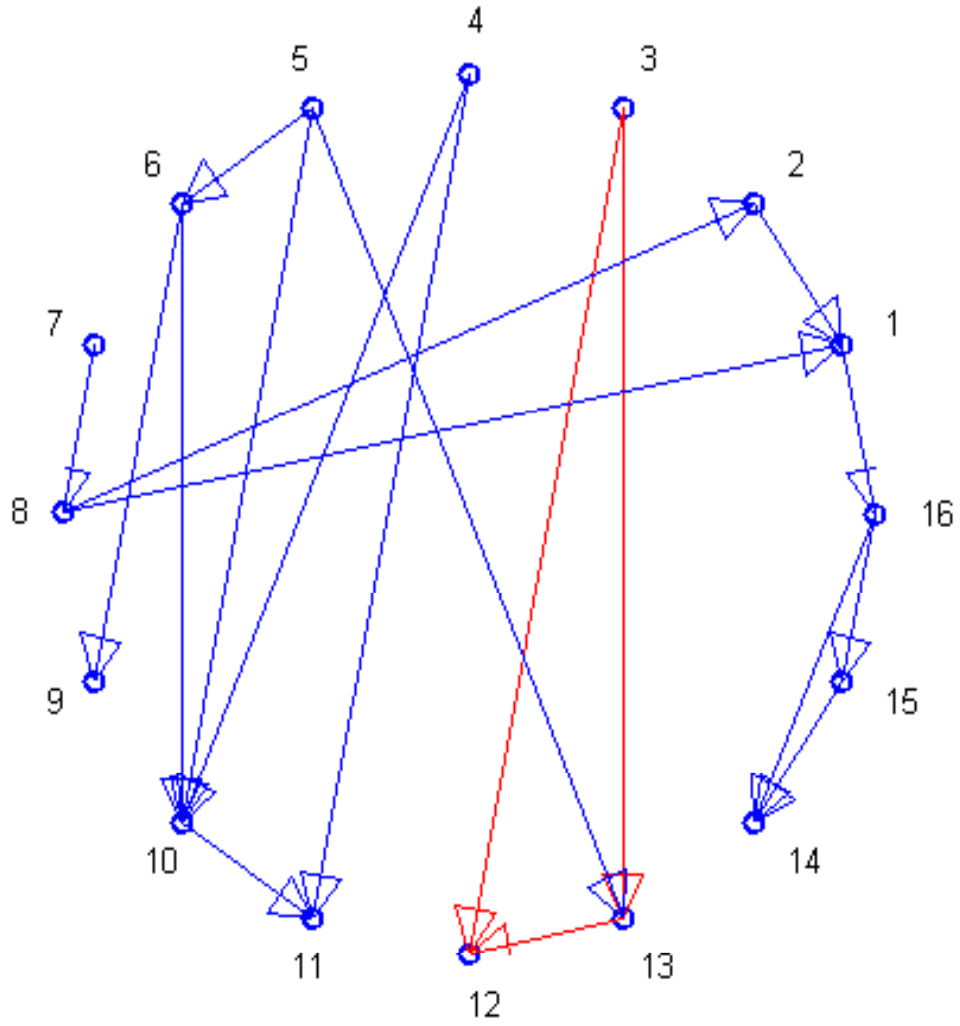
logic network **Recurring pattern** **Defined function**

Network motifs

Subgraphs which occur in the real network significantly more than in a suitable random ensemble of networks.

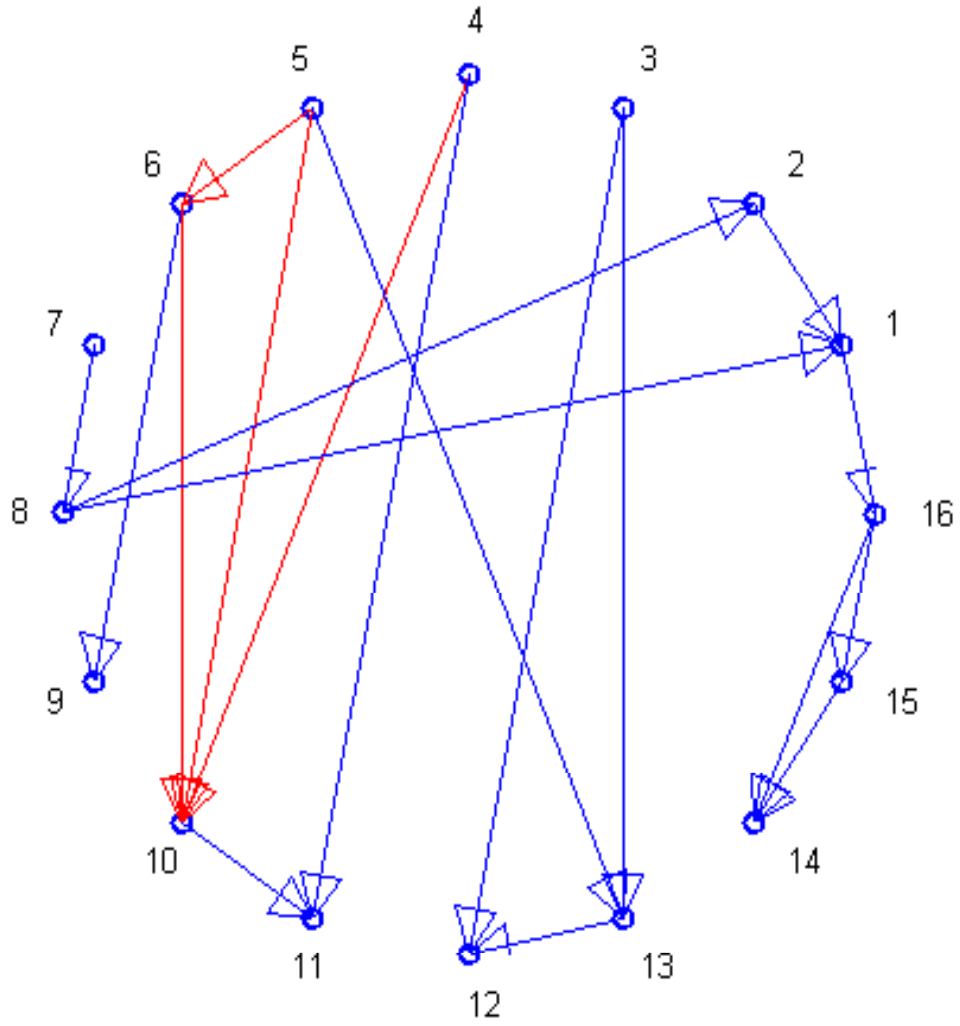
Basic terminology

3-node subgraph

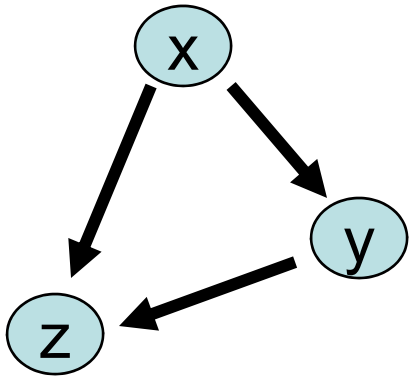


Basic terminology

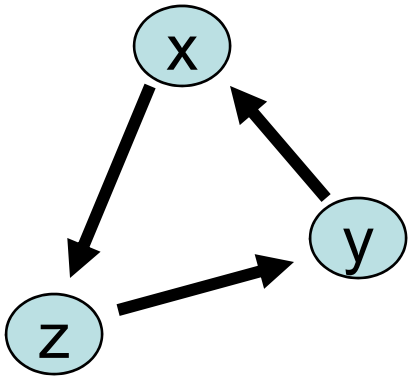
4-node subgraph



Two examples of 3-node subgraphs

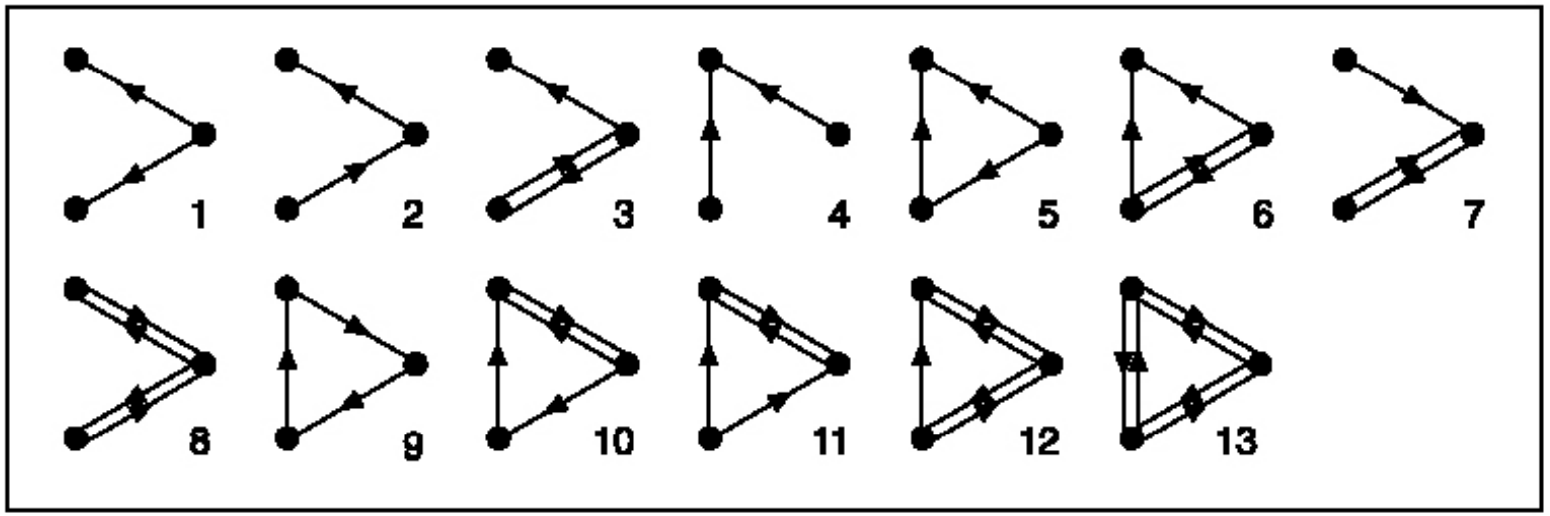


Feed-forward loop

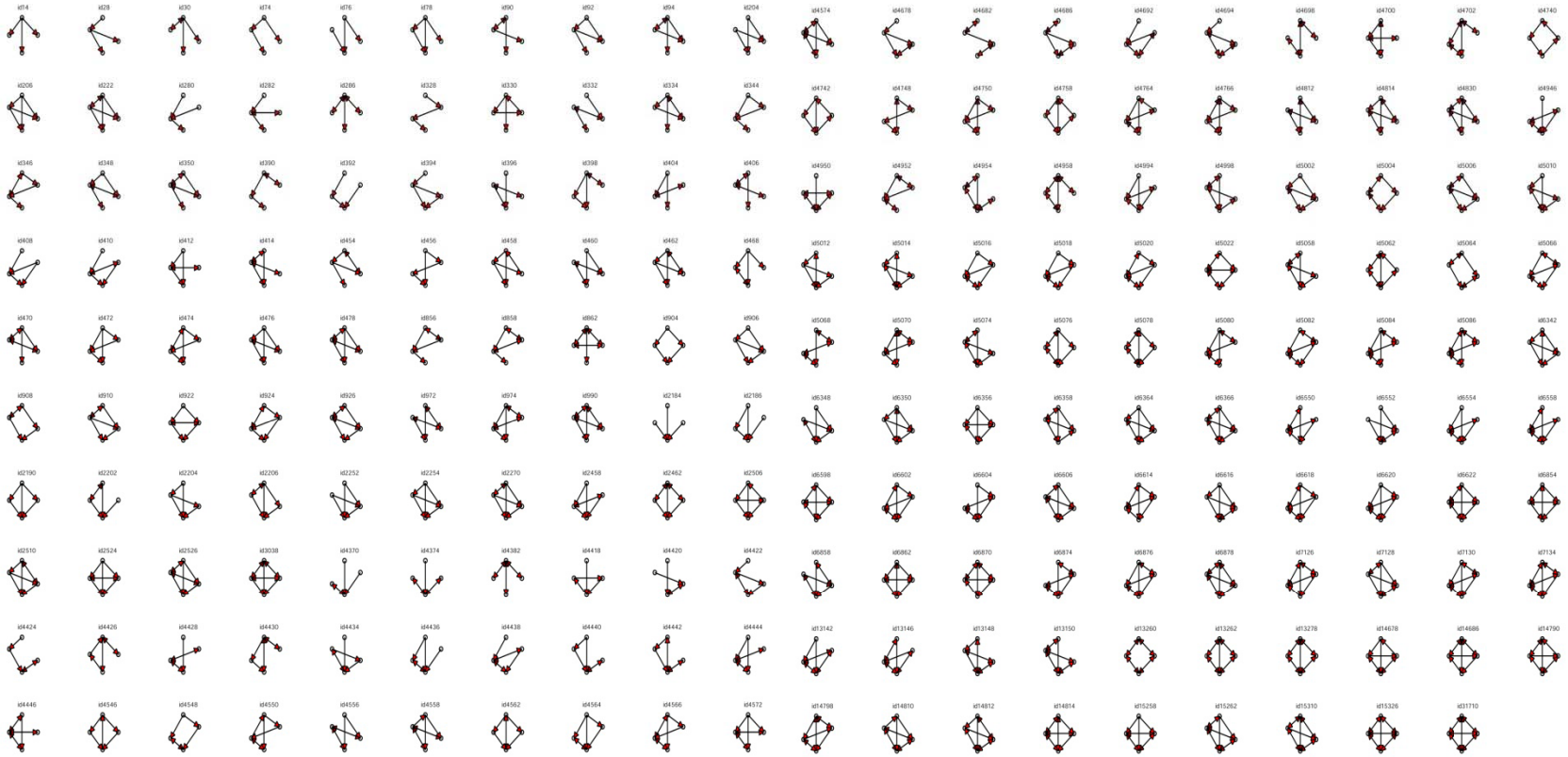


3-node feedback loop (cycle)

13 directed connected 3-node subgraphs

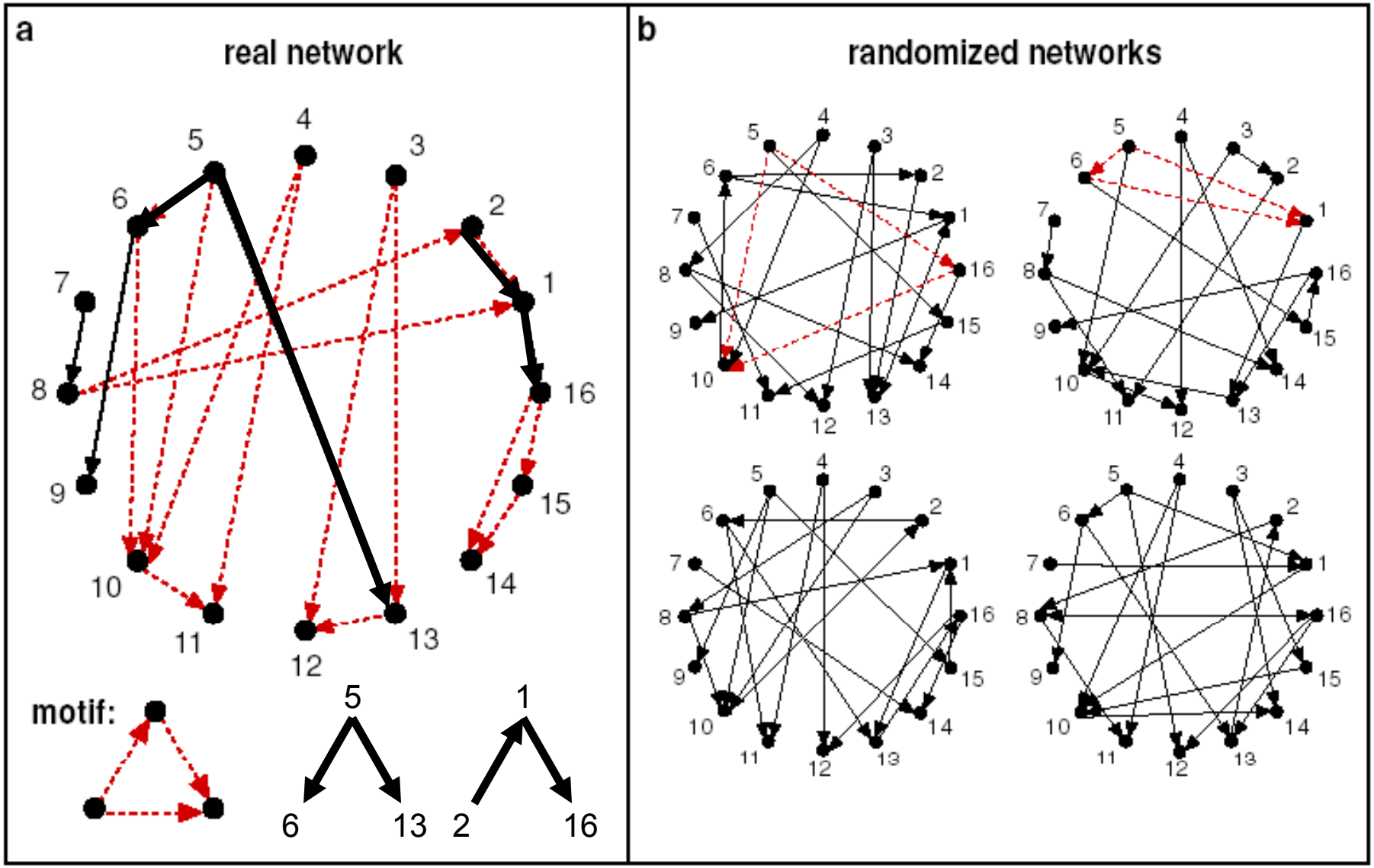


199 4-node directed connected subgraphs



And it grows pretty fast for larger subgraphs : 9364 5-node subgraphs,

1,530,843 6-node…



Real = 5

Rand=0.5±0.6

Zscore (#Standard Deviations)=7.5

Network motifs

Subgraphs which occur in the real network significantly more than in a suitable random ensemble of networks.

Algorithm :

1) count all n-node connected subgraphs in the real network.

2) Classify them into one of the possible n-node isomorphic subgraphs

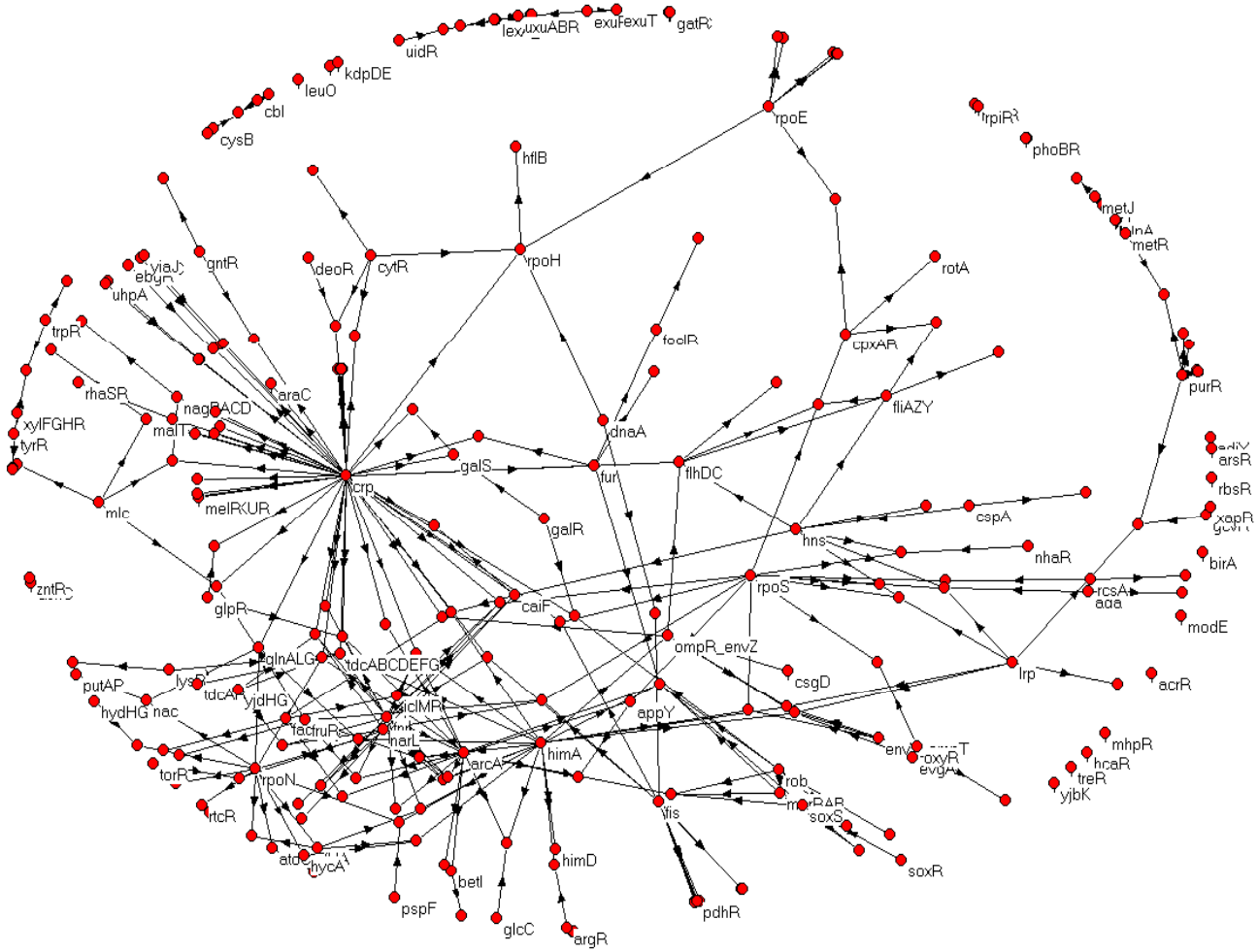
3) generate an ensemble of random networks- networks which preserve the degree sequence of the real network

4) Repeat 1) and 2) on each random network

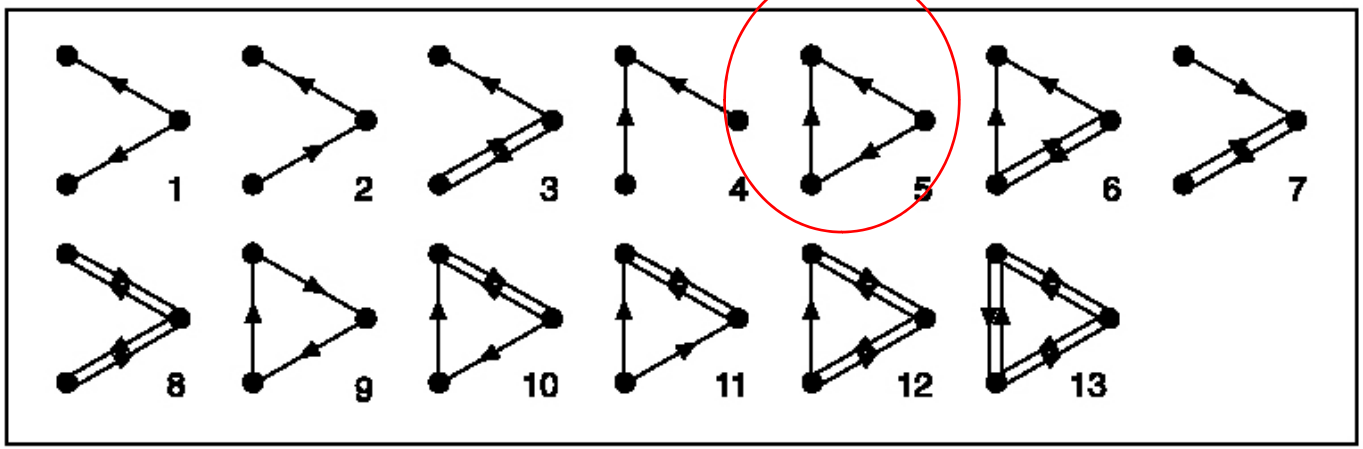
•Subgraphs with a high Z-score are denoted as network motifs.

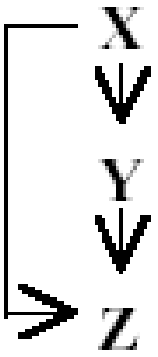
$$Z = \frac{N_{real} - N_{rand}}{\sigma_{rand}}$$

Network motifs in E. coli transcription network



Only one 3-node network motif – the feedforward loop

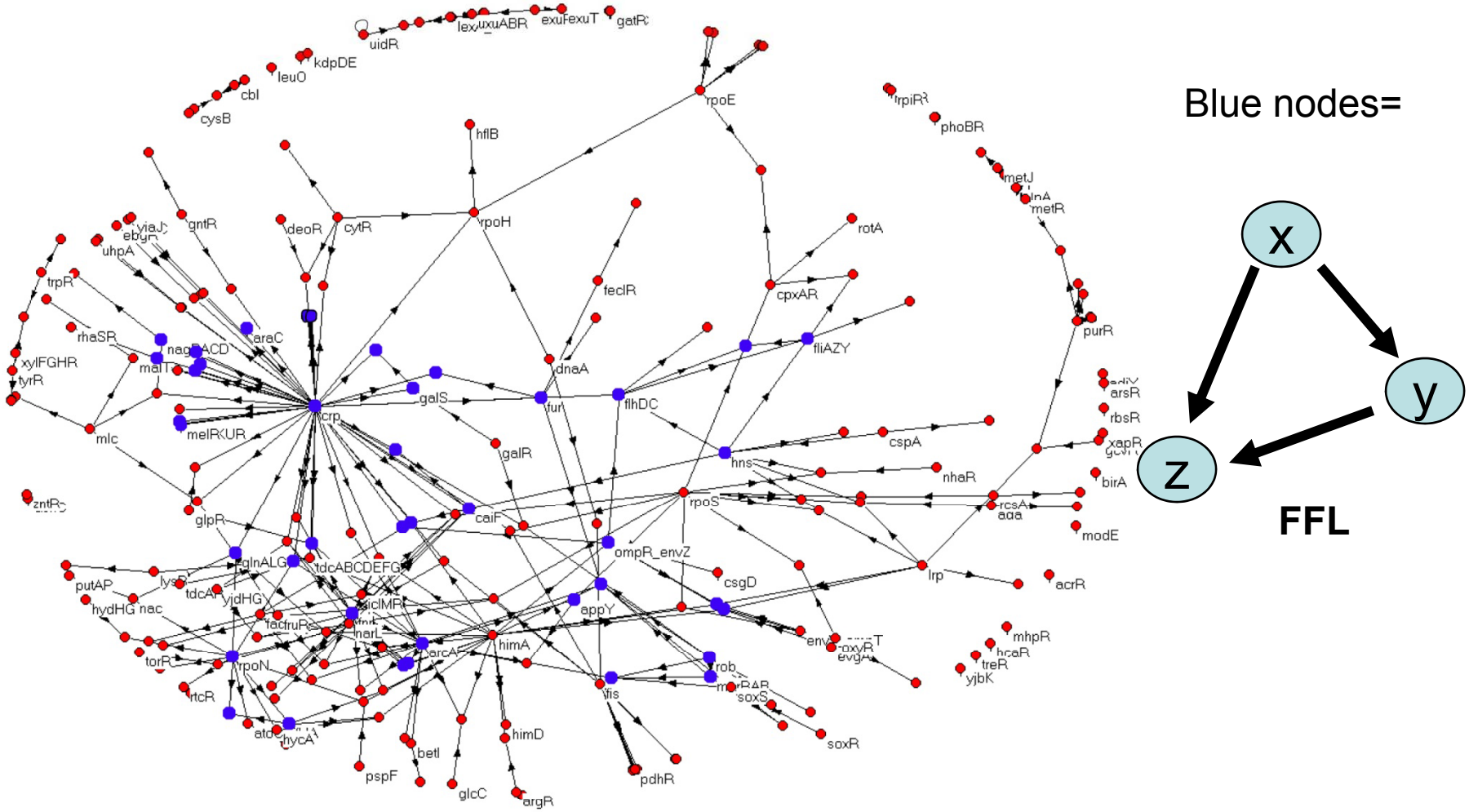




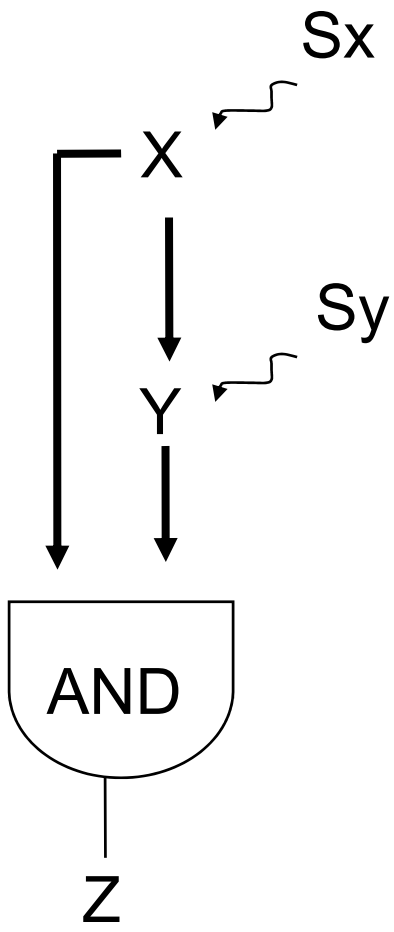
Nreal=40

Nrand=7±3

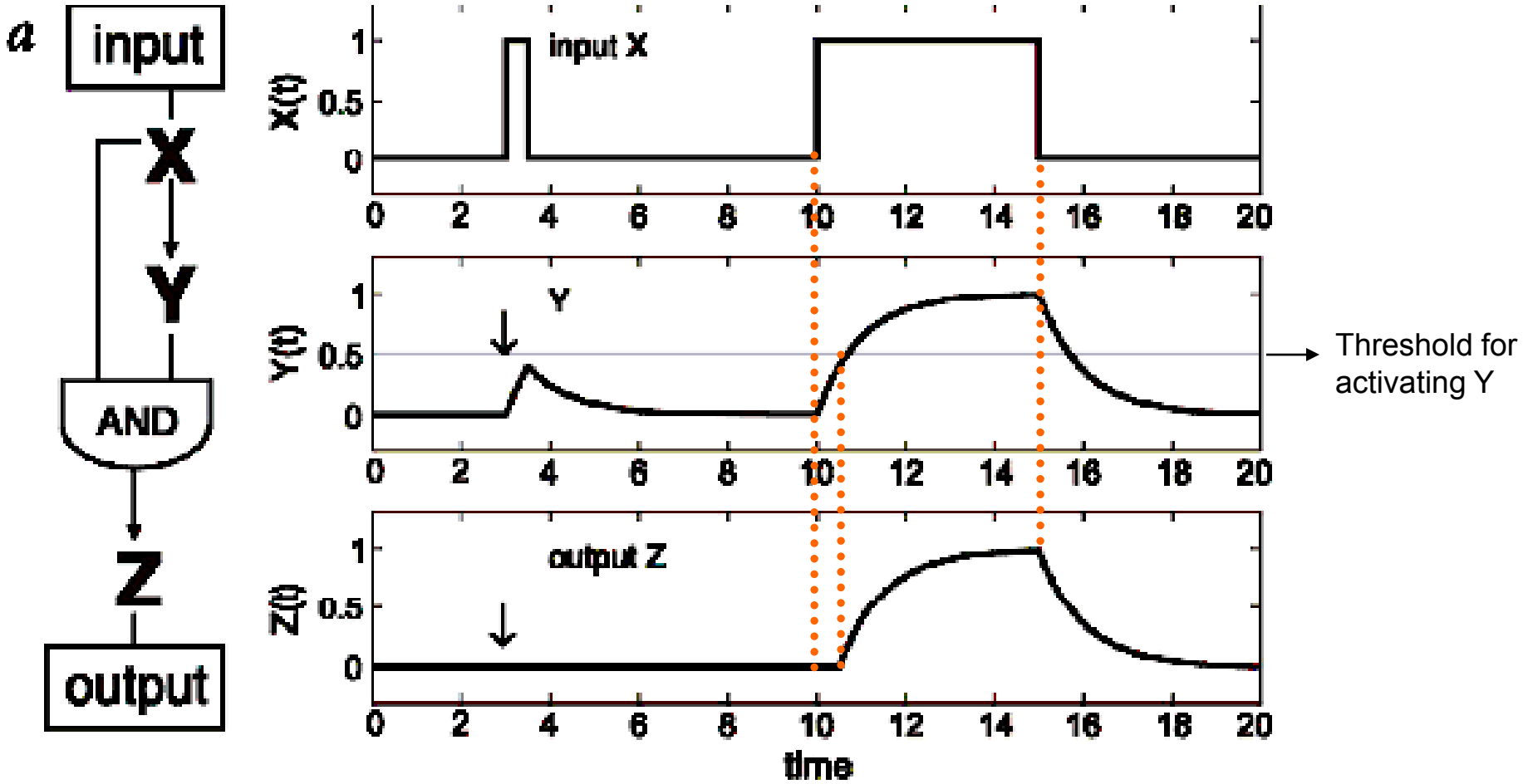
Z Score (#SD) =10


Blue nodes=
x
y
z
FFL
uidR
kdpDE
leuO
cbl
cysB
exuFexuT
gatR
hflB
rpoE
rpiR
phoBR
metJ
metR
gntR
deoR
cytR
rpoH
rotA
uhpA
trpR
fecIR
cpxAR
rhaSR
araC
purR
xylFGHR
tyrR
malT
dnaA
fliAZY
galS
fur
flhDC
arsR
rbsR
xapR
crp
mlc
melRKUR
cspA
galR
hns
nhaR
birA
rpoS
rcsA
modE
glpR
caiF
glnALG
tdcABCDEFG
ompR_envZ
csgD
lrp
acrR
putAP
lysP
tdcA
yjdHG
hydHG
nac
appY
mhpR
torR
rpoN
himA
rob
treR
yjbK
rtcR
fis
soxS
soxR
himD
betI
pspF
glcC
argR
pdhR

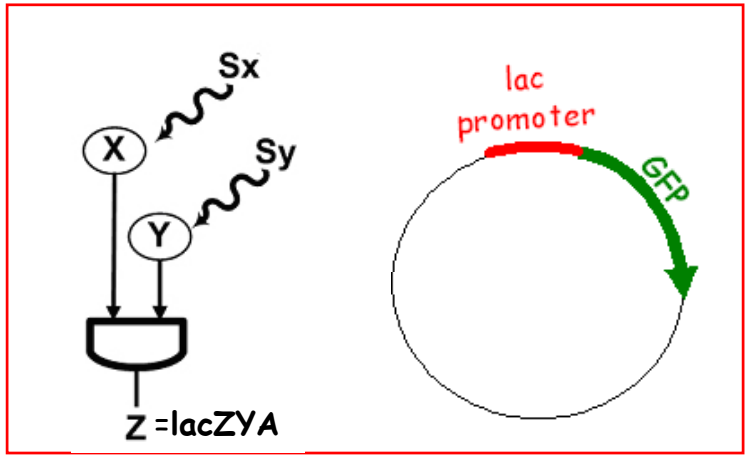
The coherent FFL circuit



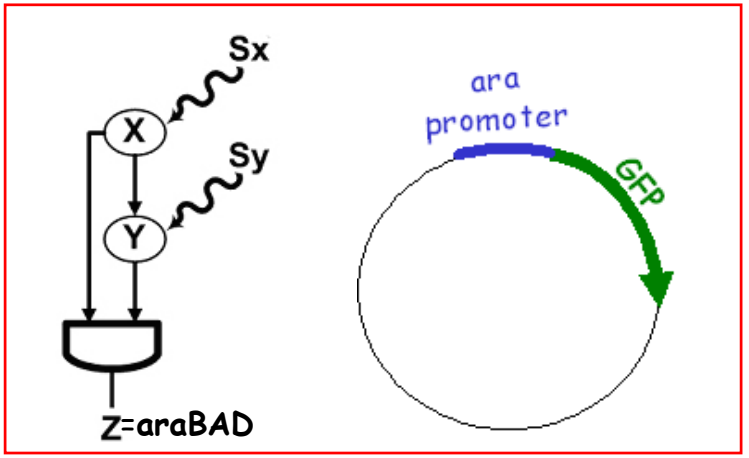
Coherent FFL – a sign sensitive filter

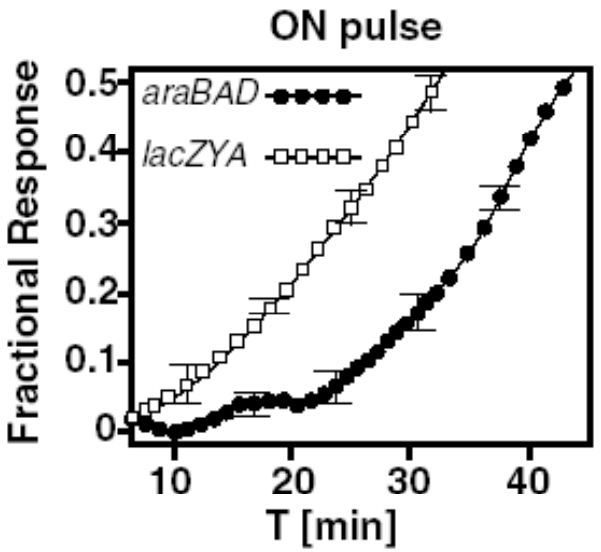


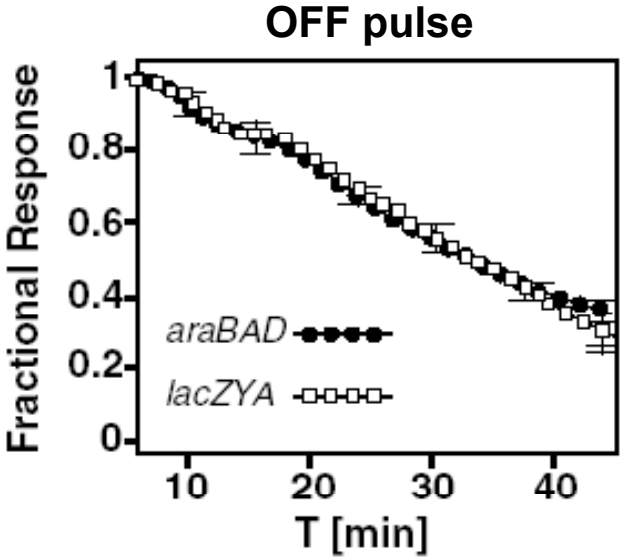
Feedforward loop is a sign-sensitive filter



Vs.

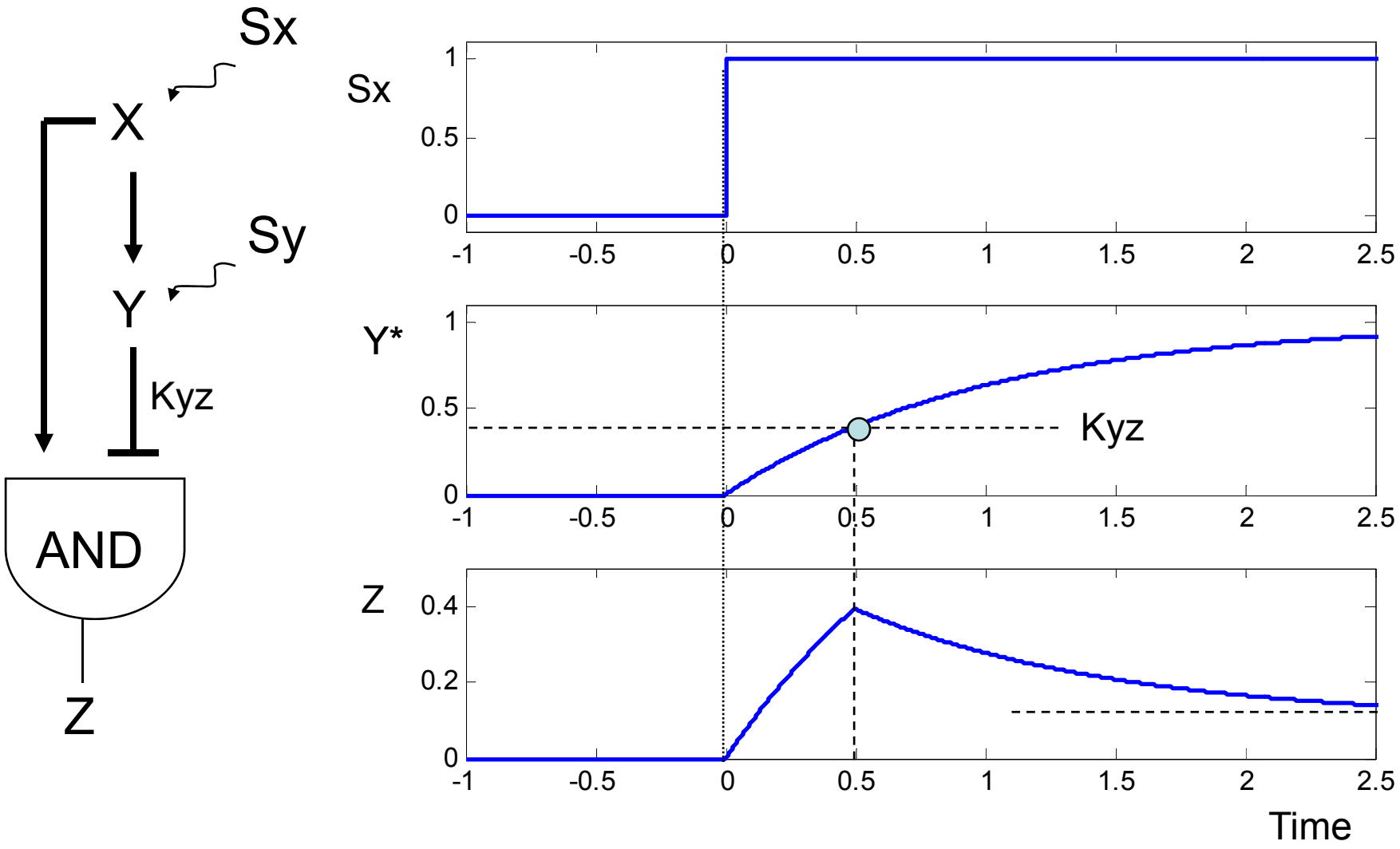




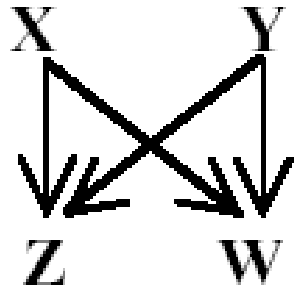


Mangan et. al. JMB

Incoherent FFL – a pulser circuit



A motif with 4 nodes : bi-fan

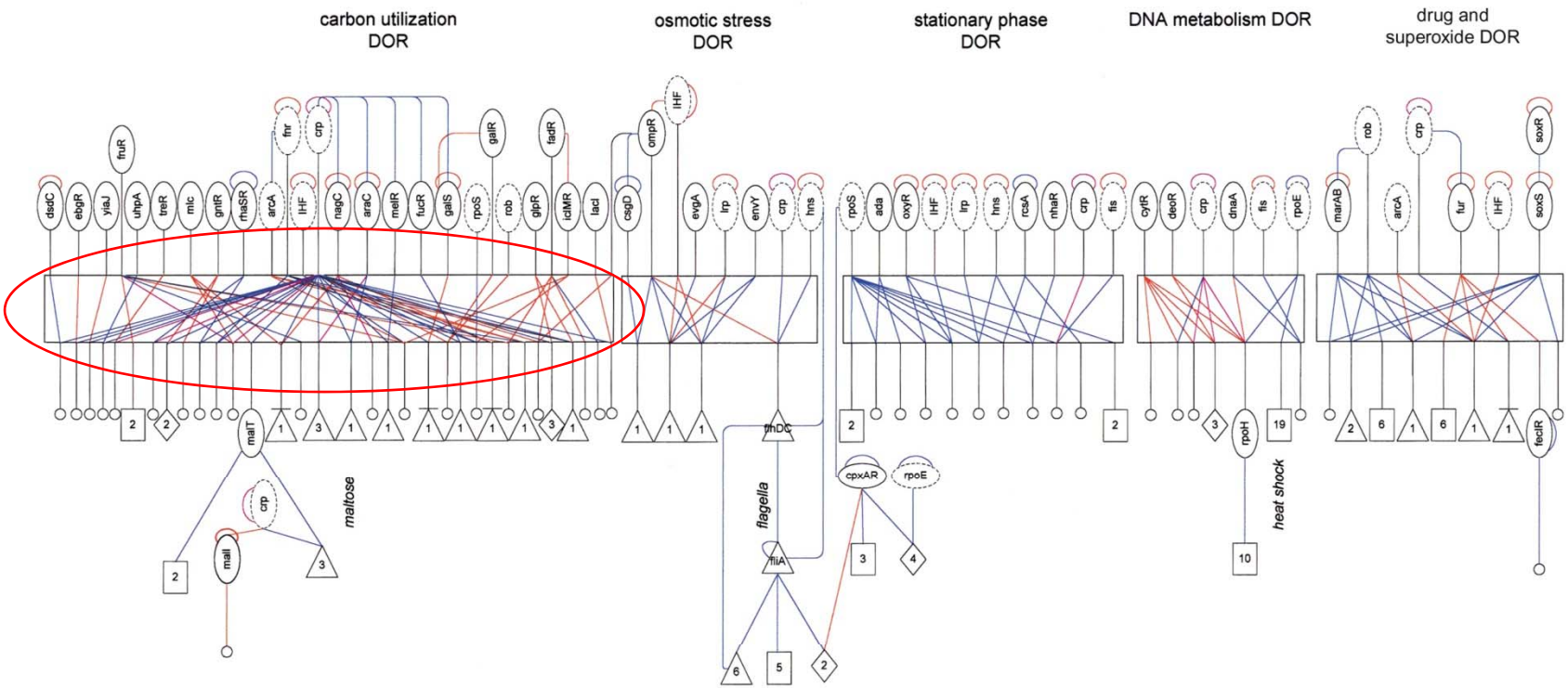


Nreal=203

Nrand=47±12

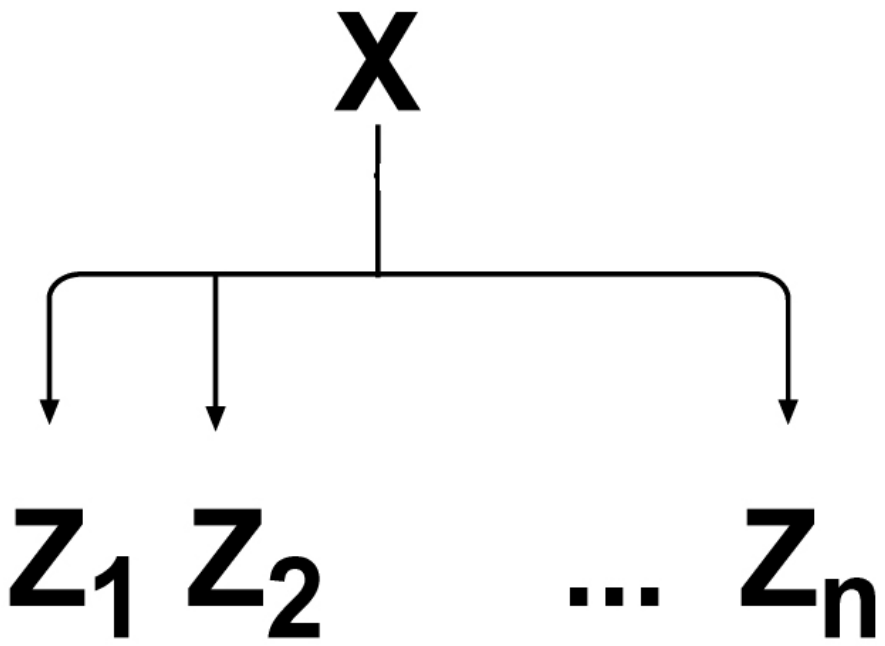
Z Score=13

bifans extend to form Dense-Overlapping-Regulons

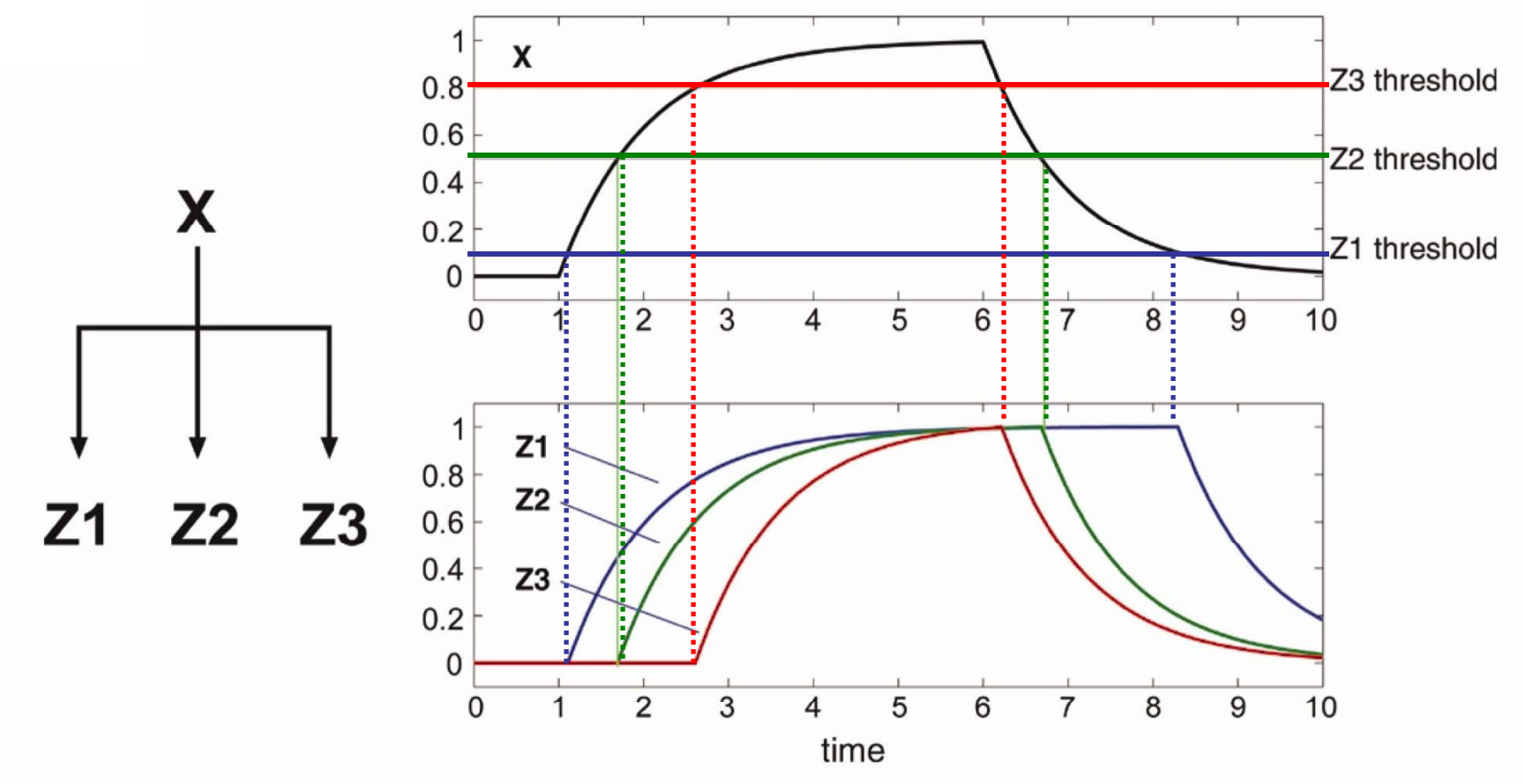


Array of gates for hard-wired decision making

Another motif : Single Input Module

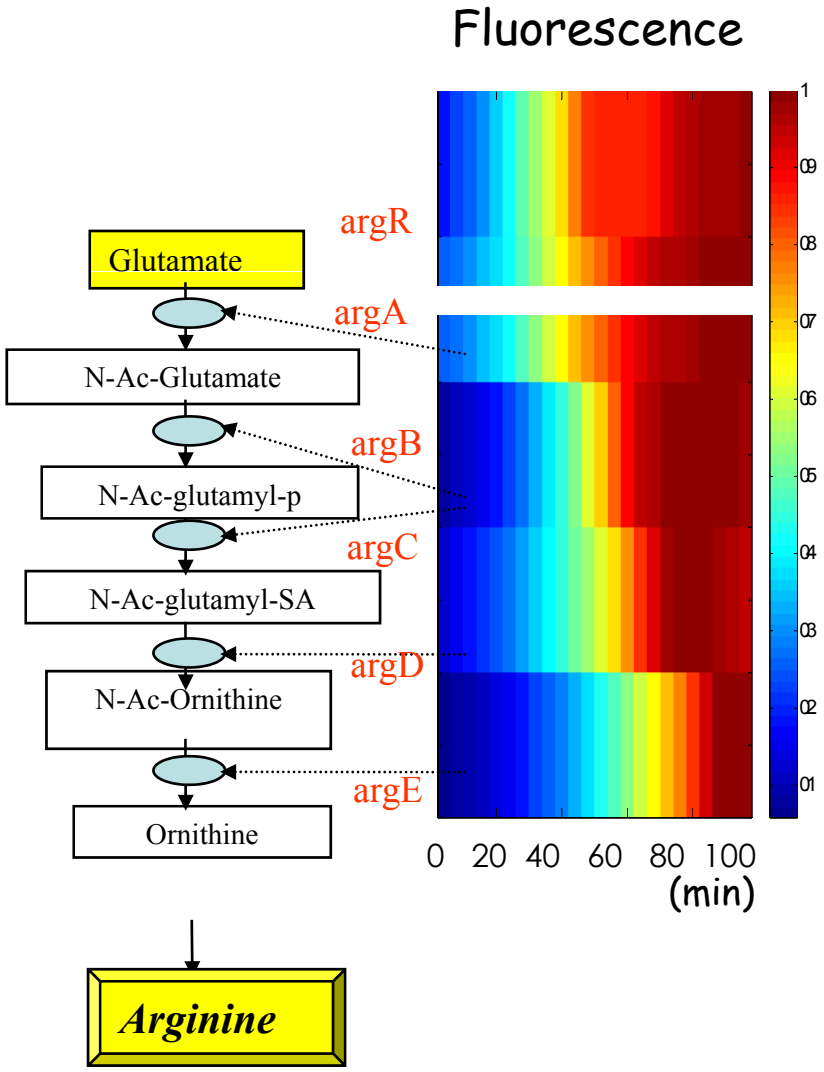


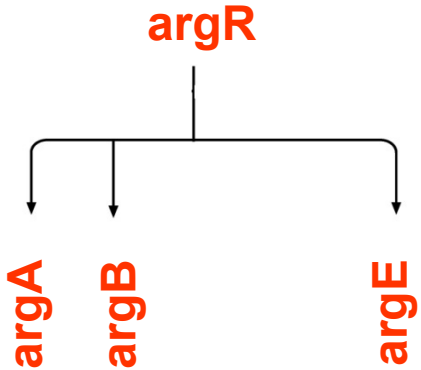
Single Input Module motifs can control timing of gene expression



Shen-Orr et. al. Nature Genetics 2002

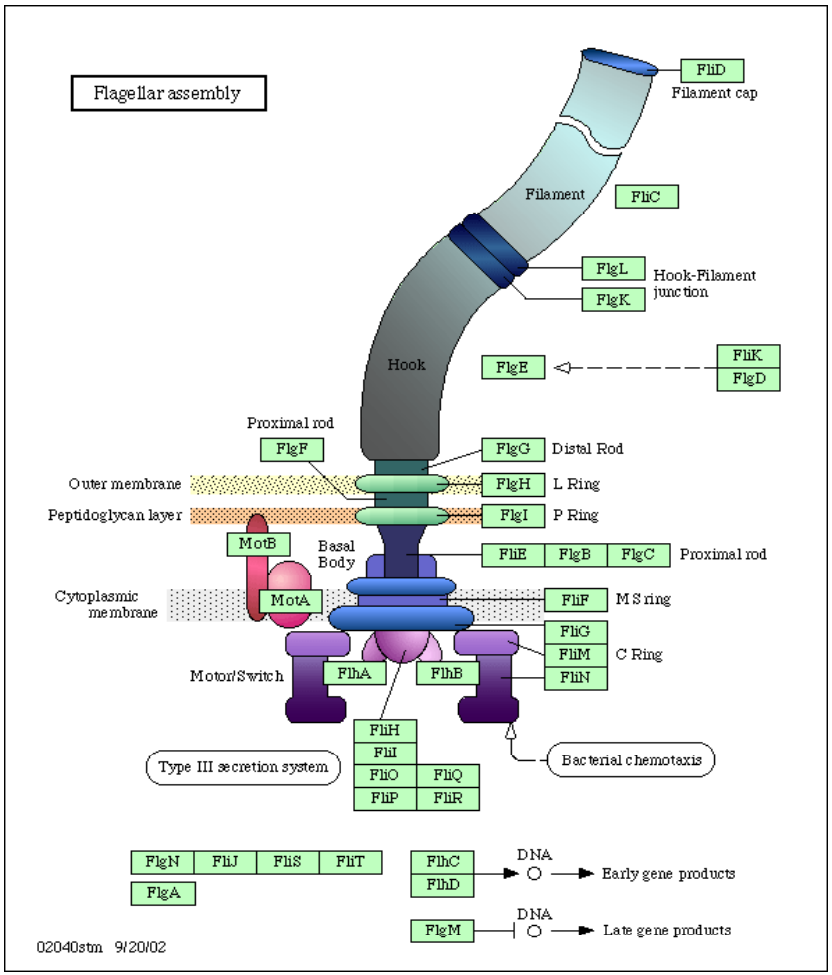
The order of gene expression matches the order of the pathway

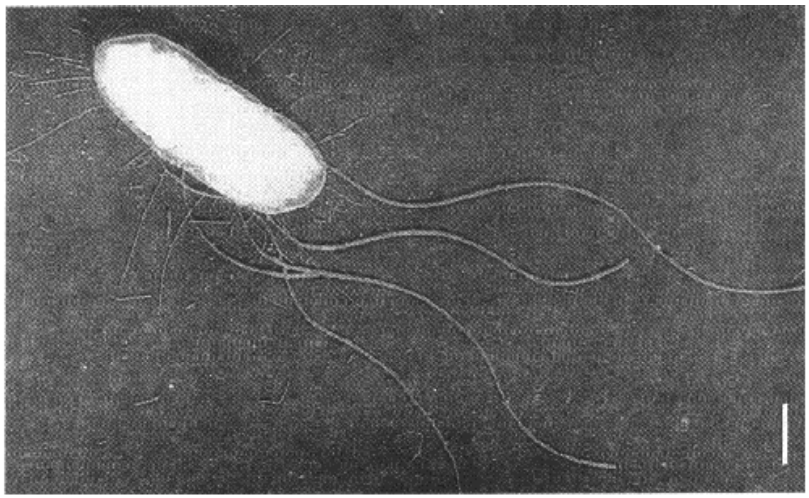




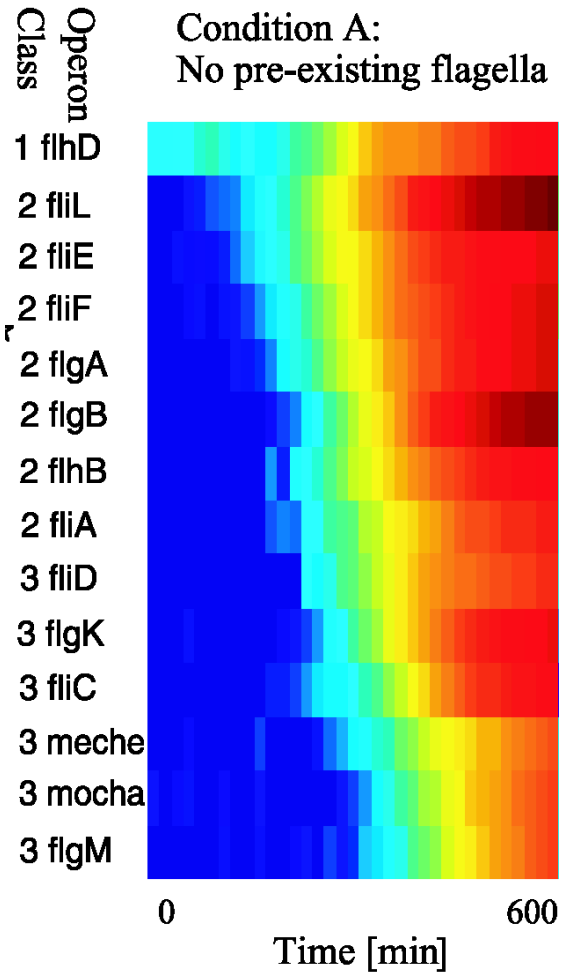
Zaslaver et. al. Nature Genetics 2004

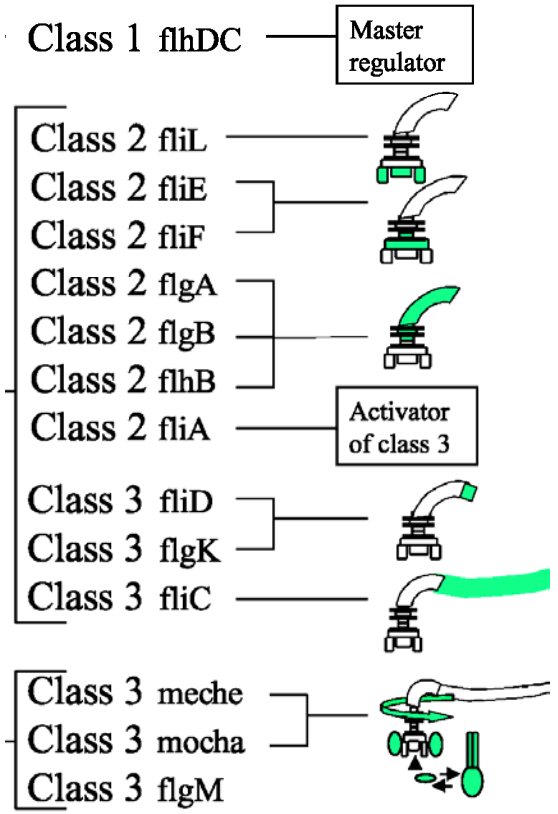
Single Input Module motif is responsible for exact timing in the flagella assembly





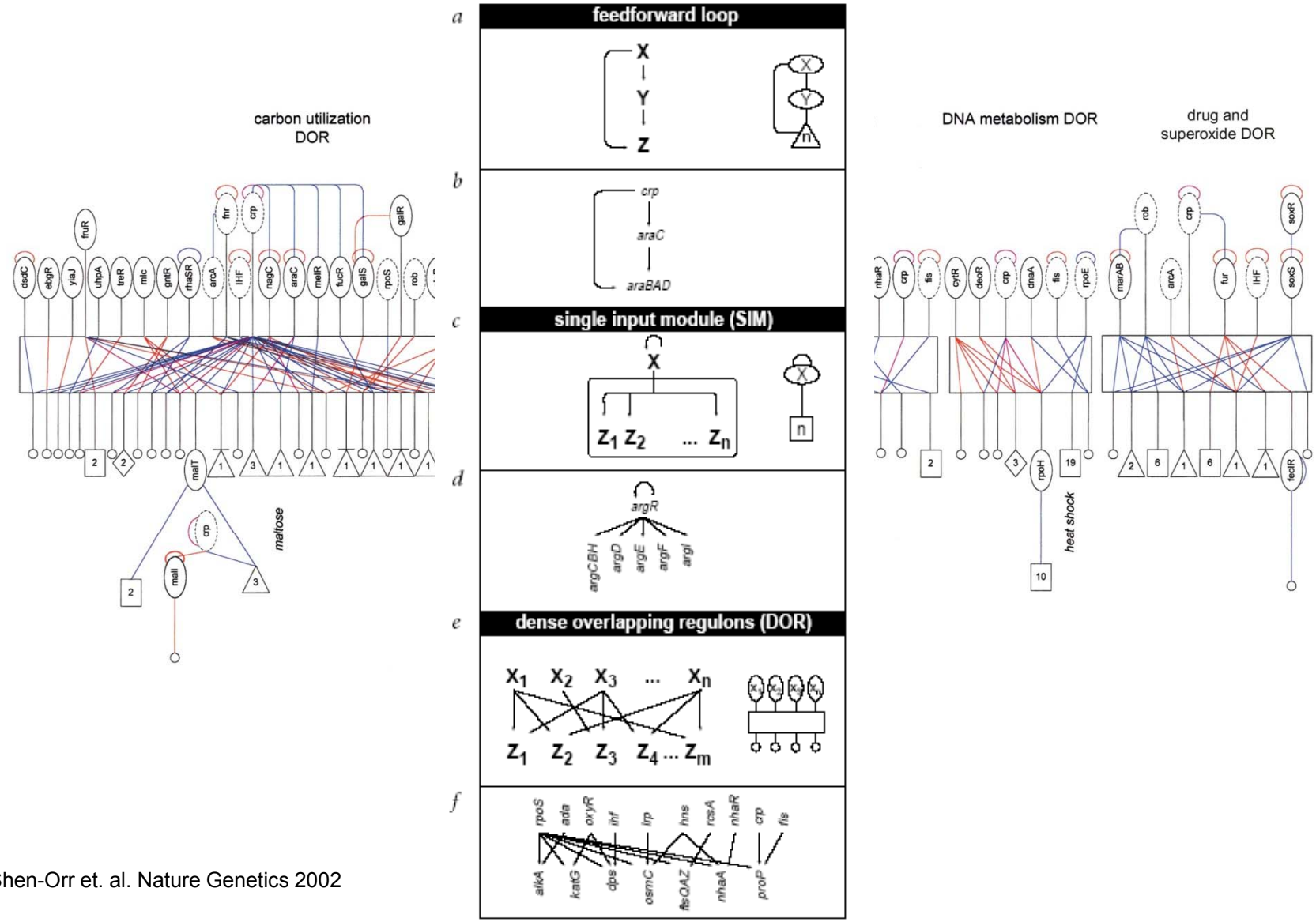
Single Input Module motif is responsible for exact timing in the flagella assembly





Kalir et. al., science,2001

The gene regulatory network of E. coli



Shen-Orr et. al. Nature Genetics 2002

Gene regulation networks can be simplified in terms of recurring building blocks

Network motifs are functional building blocks of these information processing networks.

Each motif can be studied theoretically and experimentally.

Efficient detection of larger motifs?

- The presented motif detection algorithm is exponential in the number of nodes of the motif.
- More efficient algorithms are needed to look for larger motifs in higher-order organism that have much larger gene-regulatory networks.

More information :

http://www.weizmann.ac.il/mcb/UriAlon/

Papers
mfinder – network motif detection software
Collection of complex networks