Lecture 8: Temporal programs and the global structure of transcription networks

Chap 5 of Alon

5.1 Introduction

- We will see in this chapter that sensory transcription networks are largely made of just *four families* of network motifs:
  1) auto-regulation motifs (Chap 3)
  2) FFLs with multiple outputs (Chap 4 & 5)
  3) single-input modules (SIMs), and
  4) dense overlapping regulons (DORs)

5.2 The single-input module (SIM) network motif

- **Single-input module (SIM):** a master transcription factor X controls a group of target genes Z_1, Z_2, ... , Z_n
  - Each Z_i has only one regulator (=X)
  - X usually also regulates itself
  - Regulation signs (activation vs repression) are the same for all Z_i's
  - SIM family has a free parameter n (= the number of target genes Z_i)
- In real networks, SIM is a strong network motif when compared to random networks (footnote 1, page 77/Alon)
- Fig. 5.1 (next slide)

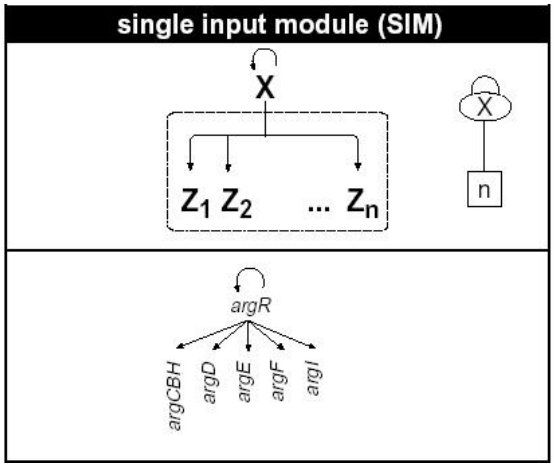
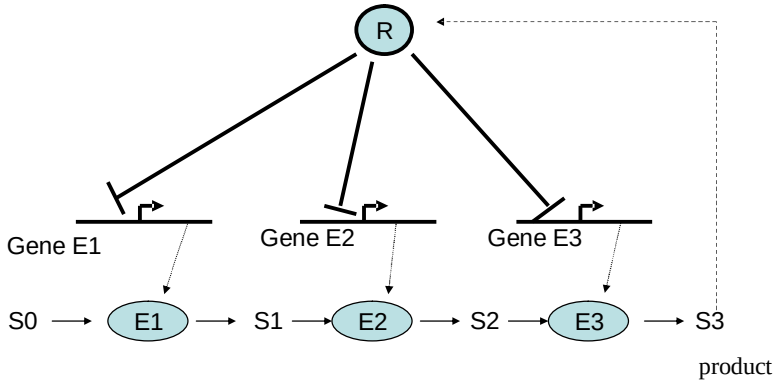


Fig 5.1 Upper panel: The single input module (SIM) network motif. Transcription factor X regulates a group of genes Z_1,.. Z_n, with no additional transcription factor inputs. X usually regulates itself. Lower panel: An example of a SIM, the arginine biosynthesis pathway (in this system, all regulations are repression).

SIM (cont.)

- The most important task of a SIM is to control a group of genes according to the signal sensed by the master regulator X
- Genes Z_i in a SIM always have a common biological function (for example, they participate in a specific metabolic pathway; see Fig 5.2 on the next slide)

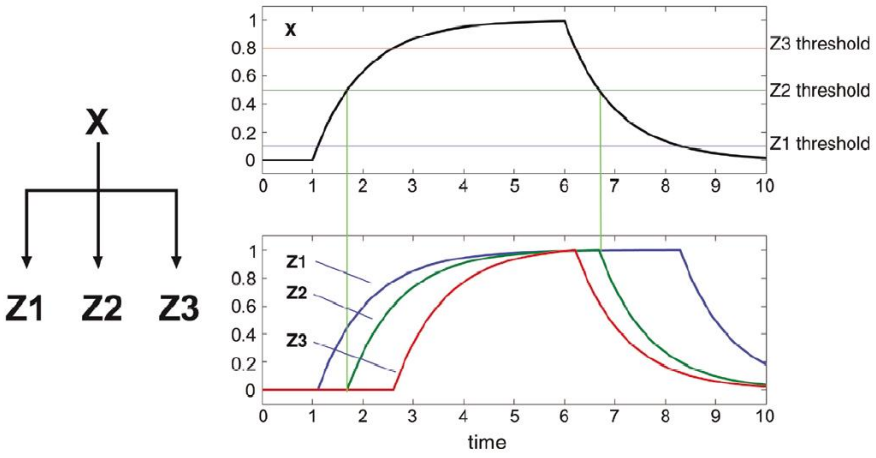
Fig 5.2 A single-input module (SIM) regulating a three-step metabolic pathway. The master repressor R represses a group of genes that encode for enzymes, E1,E2 E3 (each on a different operon). These enzymes catalyze the conversion of substrate S0 to S1 to S2 culminating in the product S3. The product S3 binds to R, and increases the probability that R is in its active state R*, in which it binds the promoters to repress the production of enzymes. This closes a negative feedback loop, where high levels of S3 lead to a reduction in its rate of production.



5.3 SIMs can generate temporal expression programs

- SIM as a temporal expression program: genes Z_i are activated one by one in a defined order, based on different activation thresholds K_i of X for each of the target genes Z_i
- Thresholds K_i depend on the specific binding sites of X
- These sites can be slightly different, resulting in different K_i for each Z_i
- When the activity level of X changes gradually in time (this is a relatively slow change for example in metabolic regulation), it crosses the thresholds K_i at different times and the genes Z_i are turned ON/OFF in the implied order
- When the level of X first goes up and then comes down, the genes Z_i are affected in the last-in-first-out (LIFO) order

Fig 5.3 The SIM can generate temporal programs of expression. As the activity of X gradually rises, it crosses the different thresholds for each target promoter in a defined order. When X activity declines, it crosses the thresholds in reverse order (last-in-first-out or LIFO order).



Source: shen-orr nature genetics 2002

Zaslaver et al Nature genetics 2004

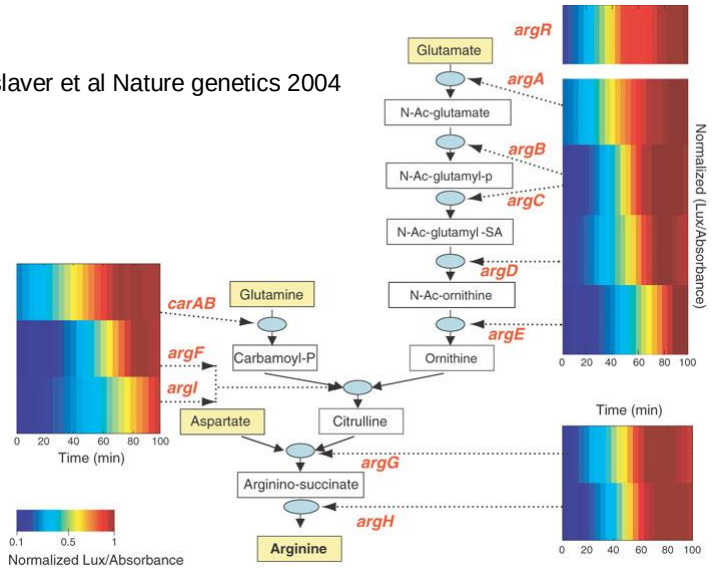


Fig 5.4 Temporal order in arginine biosynthesis system with minutes between genes. Colored bars show expression from the promoters of the different operons in the system, measured by means of a luminescent reporter gene. The position of each gene product in the pathways that produce arginine is shown.

SIMs can generate temporal expression programs (cont.)

- A unifying principle in E. coli pathways: *the earlier the protein (enzyme) functions in the pathway, the earlier its gene is activated*
- Other examples of temporal order in global cellular responses:
  - Genes timed throughout the cell cycle in bacteria and yeast
  - Genes regulated by different phases of the circadian clock that keeps track of the time of the day
  - Genes in developmental processes
- Evolution of SIMs is often convergent: the Zi's of two SIMs may be highly homologous while their regulators X are not

5.4 Topological generalizations of network motifs

- 199 possible 4-node patterns: Fig 5.5; over 9000 5-node patterns e.t.c.
- To group patterns into families with a shared functional theme, define topological generalizations of motifs
- Fig. 5.6
- The **multi-output FFL** is the only one of the generalizations of FFL in Fig 5.6 that is actually a motif (i.e., it occurs in real networks significantly often)

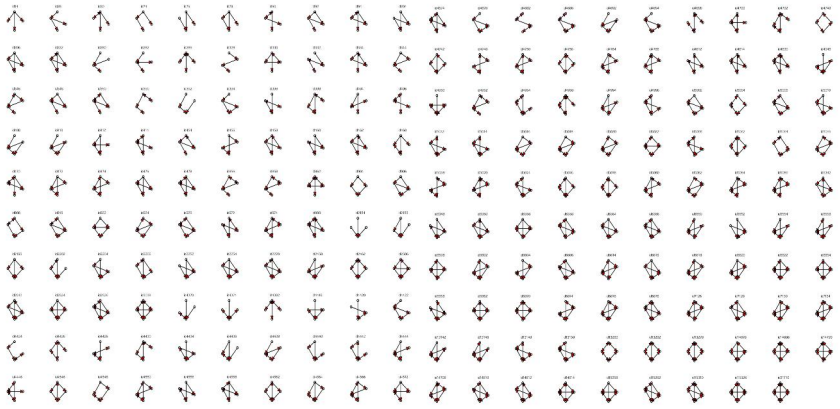
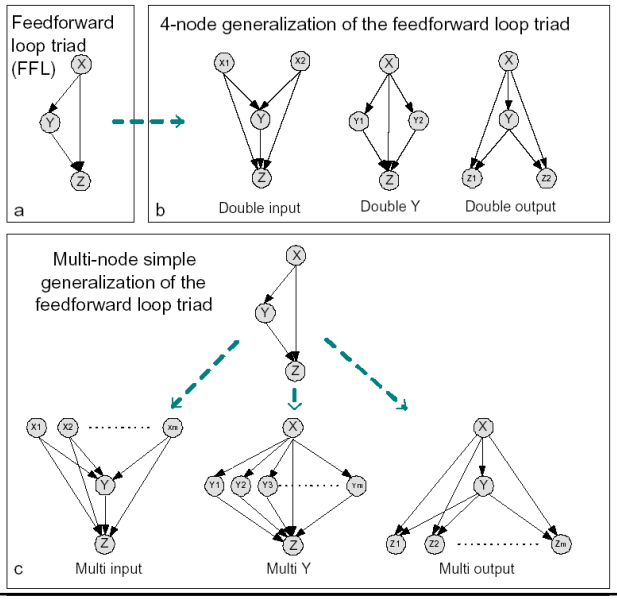


Fig 5.5 The 199 different 4-node directed connected subgraphs

Fig 5.6: Simple topological generalizations of the FFL. Each topological generalization corresponds to duplicating one of the nodes of the FFL and all of its edges. (a) The FFL; (b) Generalizations based on duplicating one node; (c) Multi-node generalizations. Source: Kashtan et al, PRE 2004.



5.5 The multi-output FFL can generate FIFO temporal order

- Example of a real multi-output FFL in the gene system that controls the production of flagella, E. coli's outboard motors:
  - When E. coli is in a comfortable environment with abundant nutrients, it divides regularly and does not try to move
  - When conditions become worse, E. coli makes a decision to grow several nanometer-size motors attached to helical flagella (propellers), which allow it to swim
- **Flagella motor** (see Fig 7.3/next slide)
  - a 50-nm device built of about 30 types of protein
  - The motor is electrical, rotating the flagellum which is a long helical filament, about 10 times longer than the cell it is attached to (length of E. coli is about 1 micron)
  - Flagella rotation (about 100 Hz) pushes the cell forward at speeds that can exceed 30 micron/sec
- The motor is put together in stages – an amazing example of biological self-assembly (see Fig 5.7)

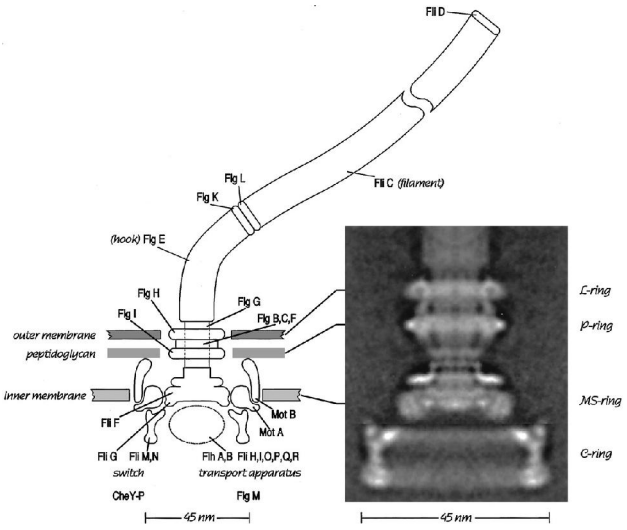
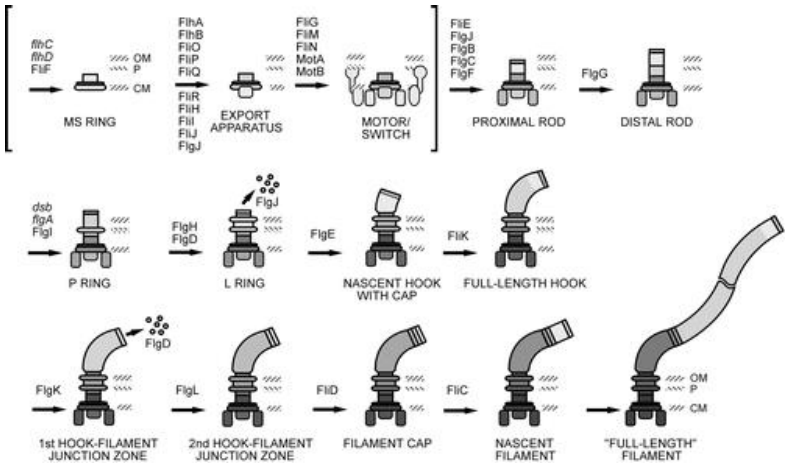


Fig 7.3 The bacteria flagella motor [source: Berg HC, Ann. Rev. Biochem 2003]

Fig 5.7: The flagella motor of E. coli and its assembly steps



info.bio.cmu.edu/.../ FlagellaMotor.html

www.aip.org/mgr/png/2002/174.htm

Annual Review of Microbiology
(2003) 100-77 :57 .Vol
(57.030502.090832.micro.annurev/10.1146:doi)
HOW BACTERIA ASSEMBLE FLAGELLA
Macnab .Robert M

Production control of the flagella motor proteins

- Proteins that build up the flagella motor are encoded by genes arranged in six operons (an operon is a group of genes transcribed on the same piece of mRNA)
- The operons are regulated by two transcription factors, both activators, X and Y
  - Master regulator X activates Y, and
  - Both jointly activate each of the six operons Z_1, Z_2, ..., Z_n
- This regulatory pattern is a multi-output FFL (see Fig 5.8)
- Each operon can be activated by X in the absence of Y, and by Y in the absence of X. Thus the input functions are similar to OR gates
- The six operons show a defined temporal order of expression (Fig 5.9)

Fig 5.8: Schematic plan of the multi-output FFL that regulates the flagella motor genes. Shown are the logic gates at each promoter, and the activation thresholds. X = flhDC, Y = fliA, Z_1 = fliL, Z_2 = fliE etc.

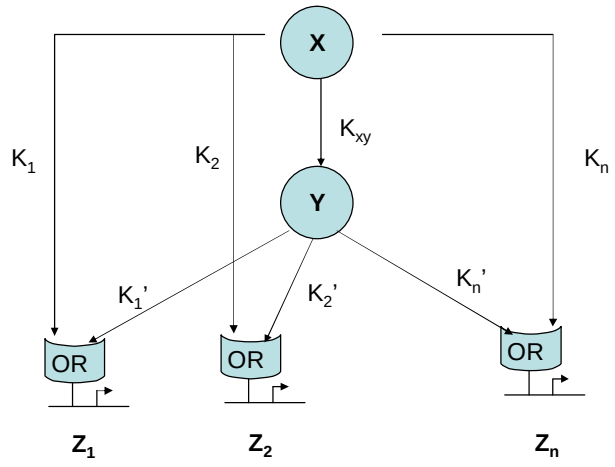
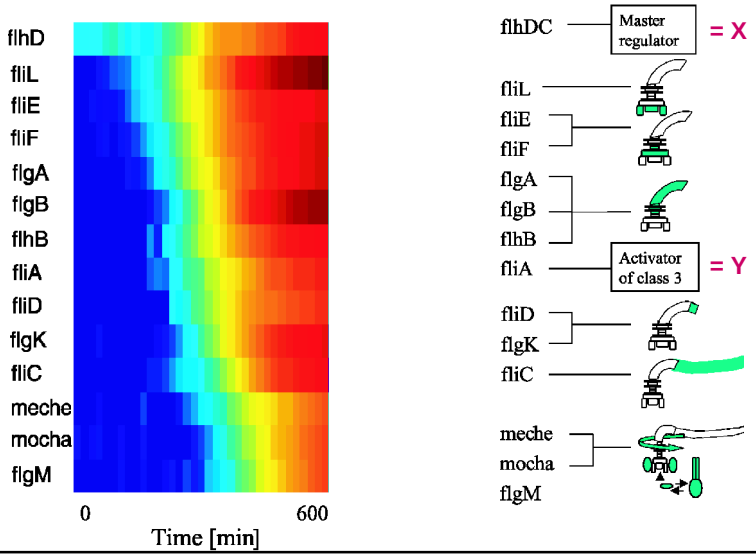


Fig 5.9: Temporal order in the flagella system of E. coli. Colored bars are the normalized expression of each promoter, where blue is low and red is high expression. Expression was measured by means of green fluorescent reporter gene. The temporal order matches the assembly order of the flagella, in which proteins are added going from the intra-cellular to the extra-cellular sides. Source: Kalir etal Science 2001



Generating FIFO order by multi-output FFL

- How can FIFO (first-in-first-out) order be generated by the multi-output FFL?
  - The activation thresholds K_{Xi} for X determine the activation order of the regulated genes
  - As the threshold K_{XY} is the smallest, Y stays on the longest time when X eventually goes down
  - Hence, as we have OR gate FFL, Y alone can keep the regulated genes Z_i still on while the level of X goes below of its activation thresholds K_{Xi} for the Z_i's
  - This means that the genes Z_i will be turned off in the order of the thresholds K_{Yi} for Y; **if this order is reversed compared to that of the thresholds K_{Xi} for X, we get the FIFO order**
  - This design was experimentally found in the flagella system (Kalir and Alon 2004)

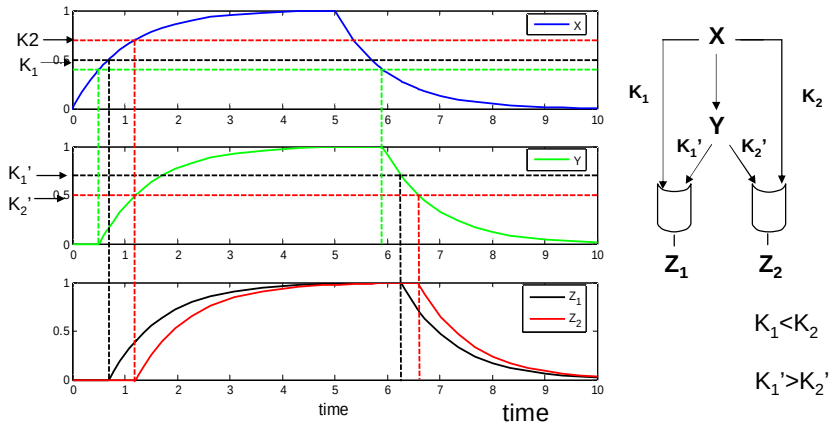


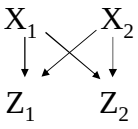
Fig 5.10: First-in-First-out order (FIFO) in the multi-Z FFL with OR-logic input functions. The output genes Z_1 and Z_2 are turned on when X crosses activation thresholds K_1 and K_2 (dashed lines). The genes are turned off when Y decays below activation thresholds K_1' and K_2'. When the order of K_1 and K_2 is opposite to that of K_1' and K_2', FIFO order is obtained.

5.6 Signal integration and combinatorial control: bi-fans and dense overlapping regulons

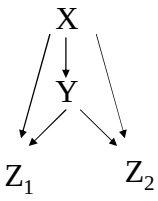
- Only two of the 199 possible 4-node petterns are significant motifs (Fig 5.11):
  - **Two-output FFL** (belongs to the family of multi-output FFLs just studied)
  - **Bi-fan**: two transcription factors X1 and X2 jointly regulate two output genes Z1 and Z2

- **Dense ovelapping regulon** or **DOR** (a generalization of bi-fan; Fig 5.12): a layer of inputs with multiple overlapping connections to a layer of outputs (Fig 5.13)
  - *Regulon* means the set of genes regulated by a given transcription factor
  - Transcription networks, such as those of E. coli and yeast, show several large DORs, each controlling tens to hundreds of genes
  - The genes in each DOR have a shared global function (stress response, nutrient metabolism,...)

Fig 5.11: The 4-node network motifs in sensory transcription networks.



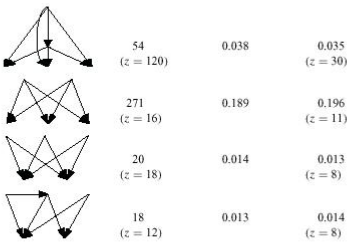
Bi-fan



Two-output

Feed-forward loop

Fig 5.12 The main five-node network motifs in the transcription network of E. coli. The bi-fan generalizes to larger patterns with a row of inputs and a row of outputs.



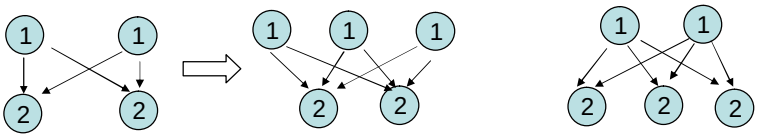
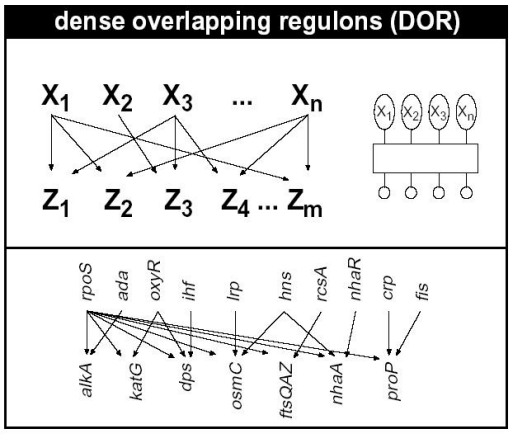


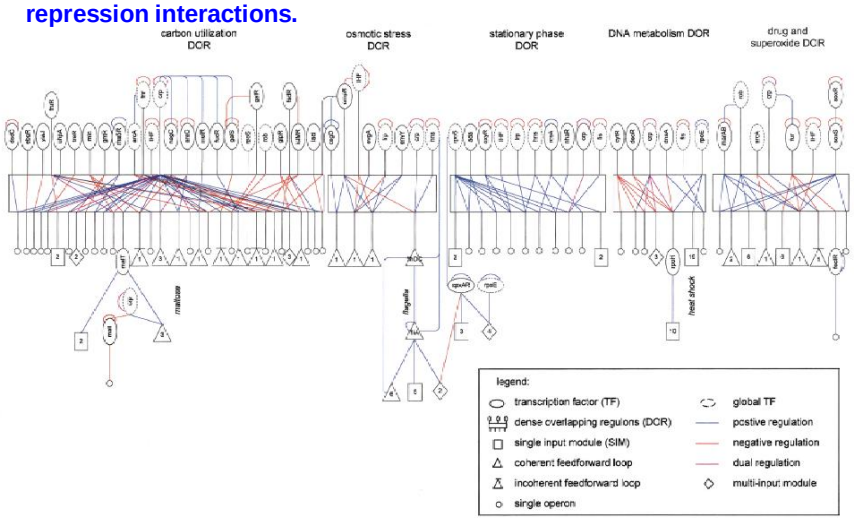
Fig 5.13 The Dense-overlapping regulons (DOR) network motif, and an example in the *E. coli* stress response and stationary phase system. Source: Shen-Orr, R Milo, S Mangan & U Alon, Nature Genetics (2002)



5.7 Network motifs and the global structure of sensory transcription networks

- To obtain a coarse-grained (easier to understand) network replace
  - each occurrence of a SIM that regulates n genes by a square marked with the number n
  - each multi-output FFL by a triangle marked with the number of output genes
  - each DOR by a rectangular box that groups its inputs, outputs, and connections
- Fig 5.14: coarse-grained network of E. coli

Fig 5.14 The global structure of part of the E. coli transcription network. Ellipses represent transcription factors that read the signals from the environment. Circles are output genes and operons. Rectangles are DORs. Triangles are outputs of single- or multi-output FFLs. Squares are outputs of SIMs. Blue and red lines correspond to activation and repression interactions.



Network motifs and the global structure (cont.)

- The four types of motifs (auto-regulation, FFLs, SIMs, and DORs) cover virtually all genes in the organisms studied so far, including the known part of the sensory networks of bacteria, yeast, worms, fruit flies, and humans
- A striking feature of the organization of sensory networks: relative absence of long cascades of transcription interactions, X → Y → Z → , etc. Most genes are regulated just one step away from their transcription factor inputs
- Why are long cascades rare in sensory networks?
  - A possible reason is response time constraints: long cascades would typically be far too slow for sensory transcription networks that need to respond quickly to environmental stresses and nutrients
- Other networks (developmental transcription networks, signal transduction networks whose components are fast) do have long cascades (see Chap 6/Alon)
- The four motif families (auto-regulation, FFLs, SIMs, and DORs) are summarized in Fig 5.15

Fig 5.15 Network Motifs in sensory transcription networks

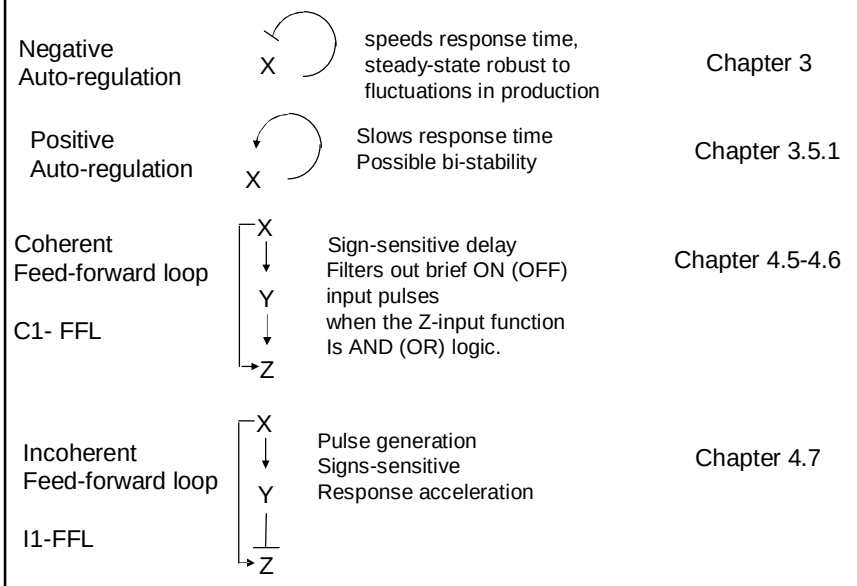


Fig 5.15 cont: Network Motifs in Sensory transcription networks