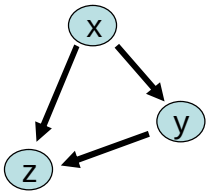
Lecture 6: The feed-forward loop (FFL) network motif

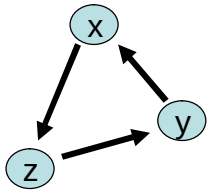
Chapter 4 of Alon

4.1 Introduction



Feed-forward loop
(**FFL**)

a=1



3-node feedback loop
(**3Loop**)

a=3

Fig 4.1a The feed-forward loop (FFL) and the feedback-loop (3Loop), two examples of subgraphs with three nodes

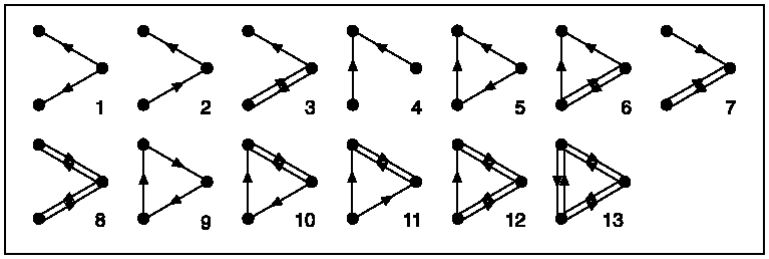


Fig 4.1b : The thirteen connected three-node directed subgraphs.
The feed-forward loop is subgraph 5, and the feedback loop is subgraph 9.

Patterns with 3 nodes

- 13 possible 3-node patterns
- Turns out that only one of them, the feed-forward loop (FFL) is a network motif (i.e., it occurs unexpectedly often in a transcription network)
- 8 possible FFL types (as there are 8 combinations of the three signs +,- on the edges)
- 2 of them appear in larger amounts
  - Why?
  - Why the others are not common?

4.2 The number of appearances of a subgraph in random networks

- Let an ER network have N nodes and g edges
  => the probability that a fixed edge is present is $p = E/N^2$
- *Important:* most biological networks are sparse: $p << 1$
  - Our example network of E. coli: N=400, E=500 => p=0.003

- Problem: How many times a given subgraph G appears in a random ER-network?
- *Solution*
  - Assume: G has n nodes and g edges
  - Let $\langle N_G \rangle$ denote the average (expected, mean) number of occurrences of G in ER-network with N nodes and E edges. Then

$$\langle N_G \rangle \approx a^{-1} N^n p^g$$

where a is the number of permutations of the n nodes of G that give an isomorphic subgraph (a=1 for FFL; a=3 for 3-node feedback loop)

Number of appearances (cont.)

- Mean connectivity (average number of edges per node):
  $\lambda = E / N$
- Hence $\langle N_G \rangle$ can be written
  $\langle N_G \rangle \approx a^{-1} \lambda^g N^{n-g}$
- i.e., the larger is the connectivity, the more there are subgraphs G
- Scaling relation: dependence of $\langle N_G \rangle$ on N:
  $\langle N_G \rangle \sim N^{n-g}$
- Examples:
  - V-shaped G: n=3, g=2 => $\langle N_{V\text{-shaped}} \rangle \sim N^{n-g} = N$ => V-shaped G is common
  - Fully connected clique: n=3, g=6 => $\langle N_{3\text{-clique}} \rangle \sim N^{-3}$ => 3-clique is rare in large ER-networks
  - FFL: n=3, g=3 => $\langle N_{FFL} \rangle \sim \lambda^3 N^0 = \lambda^3$
  - 3-node feedback loop: n=3, g=3 => $\langle N_{3loop} \rangle \sim 1/3 \cdot \lambda^3 N^0 = 1/3 \cdot \lambda^3$
- This is remarkable: *Average number of FFLs and 3Loops are constant in ER networks and do not increase with network size* (when the connectivity stays constant)

4.3 The feed-forward loop is a network motif

- Fig 4.2: $N_{FFL} = 42$, $N_{3loop} = 0$

- $\langle N_{FFL} \rangle_{rand} \sim \lambda^3 \sim 1.23 \sim 1.7$
- $\langle N_{3loop} \rangle_{rand} \sim \lambda^3/3 \sim 0.6$

=> FFL is a strong network motif (Z=31 for ER networks; Z=7 for degree-preserving random networks (pp 45-46 in texbook))

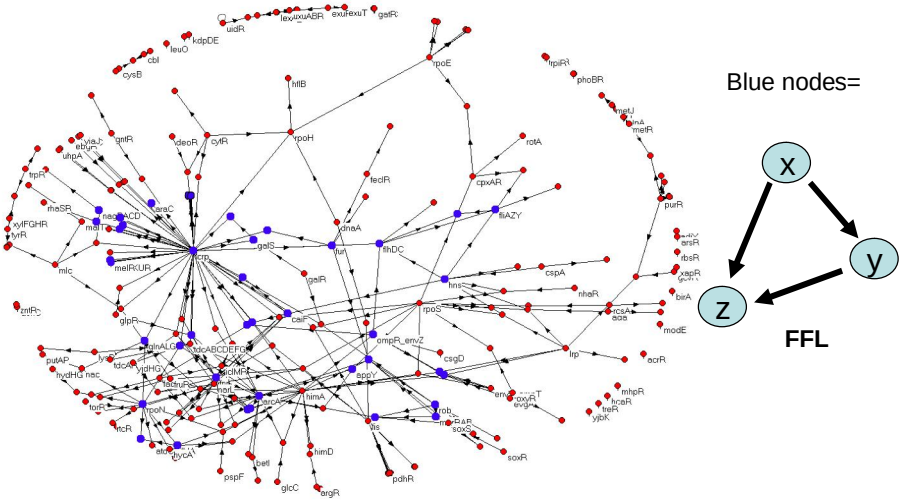
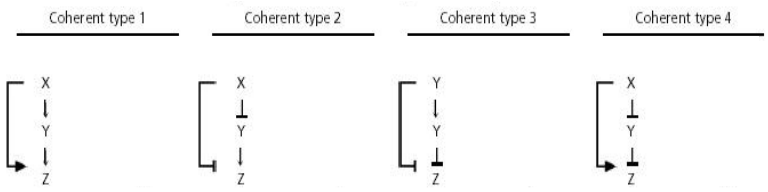


Fig 4.2 Feed-forward loops in the E. coli transcription network. Blue nodes participate in FFLs.

4.4 The structure of the feed-forward loop gene circuit

COHERENT FFL



INCOHERENT FFL

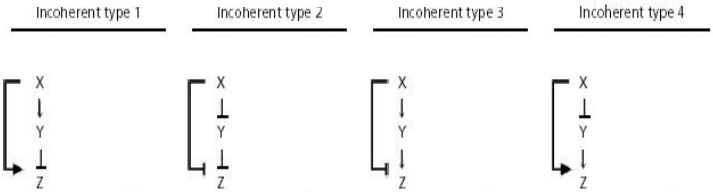


Fig 4.3 The 8 sign combinations (types) of FFLs. Arrows denote activation and --| symbols denote repression.

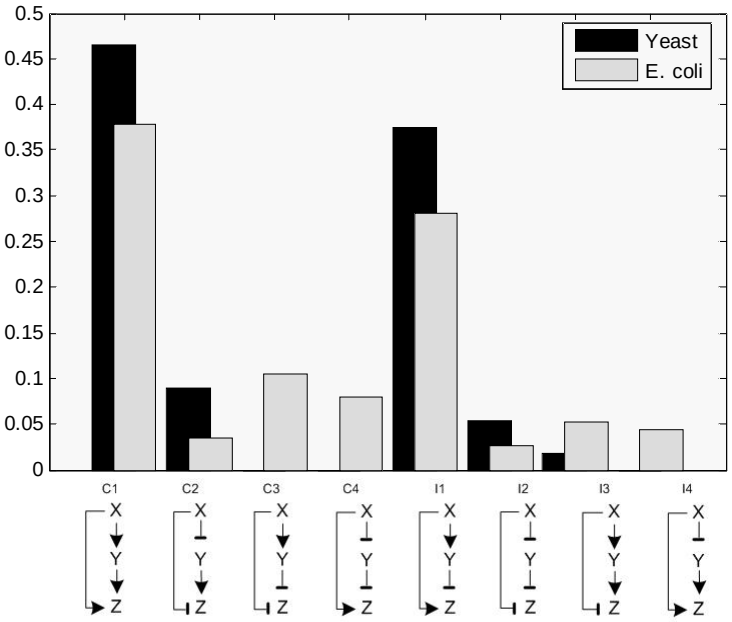


Fig 4.4 Relative number of the eight FFL types in the transcription network of yeast and E. coli. FFL types are marked C and I for coherent and incoherent. The E. coli network is based on the Ecocyc and Regulon DB databases (about twice as many edges as in the network of Fig 2.3).

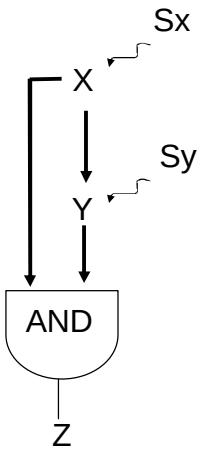


Fig 4.5a The coherent type-1 FFL (C1-FFL) with an AND input function: Transcription factor X activates the gene encoding transcription factor Y, and both X and Y jointly activate gene Z. The two input signals are Sx and Sy. An input-function integrates the effects of X and Y at the Z promoter (an AND-gate in this figure).

4.5 Dynamics of the coherent type-1 FFL (C1-FFL) with AND logic

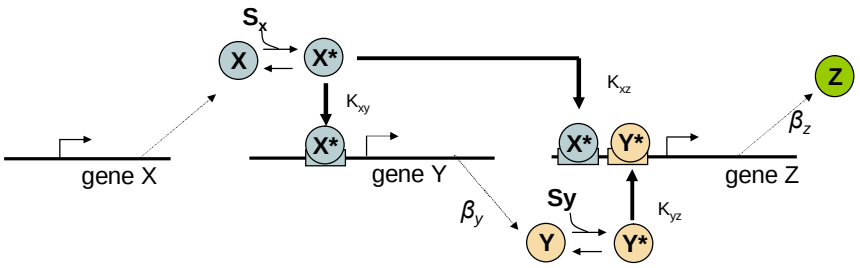


Fig 4.6 The molecular interactions in the type-1 coherent FFL.
The protein X is activated by signal S_x, which causes it to assume the active conformation X*. It then binds its sites in the promoters of genes Y and Z. As a result, protein Y accumulates, and, in the presence of its signal S_y, is active Y*. When Y* concentration crosses the activation threshold K_{yz}, Y* binds the promoter of gene Z. Protein Z is produced when both X* and Y* bind the promoter of gene Z (an AND input function).

4.6 C1-FFL is a sign-sensitive delay element

- ON step: S_X suddenly appears
- OFF step: S_X suddenly disapperas
- Assume that Y is always present in its active form: Y* = Y

Delay following an ON step of S_X

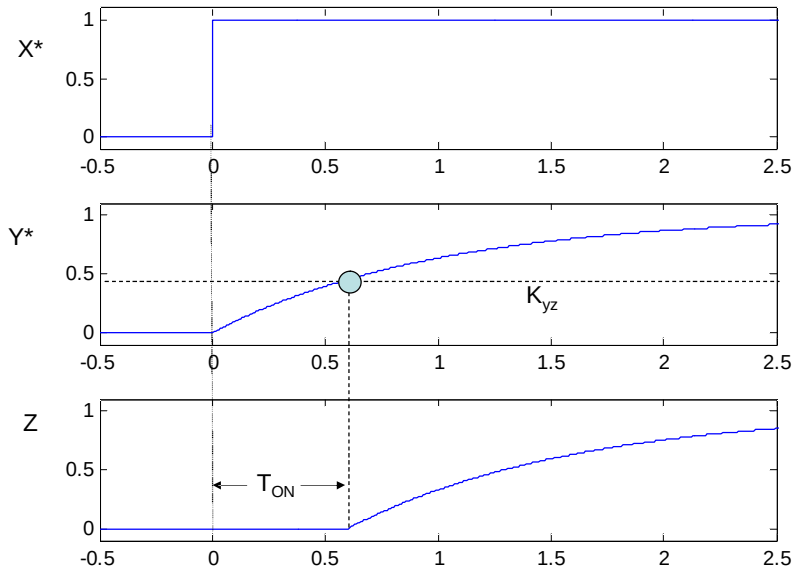


Fig 4.7: Dynamics of the coherent type-1 FFL with AND logic following an ON step of S_x at time t=0. The activation threshold of Z by Y is K_{yz} (dashed line). The production and degradation rates are $\alpha_y=\alpha_z=1$, $\beta_y=\beta_z=1$. The delay in Z production is T_{ON}.

Delay following an ON step of S_X (cont.)

$$Y^*(t) = Y_{st}(1 - e^{-\alpha(Y)t})$$

$$Y_{st} = \beta_Y / \alpha_Y$$

- Z begins to be expressed after a delay T_{ON} (= time for Y to reach K_{YZ})

  $$\Rightarrow\ Y^*(T_{ON}) = Y_{st}(1 - e^{-\alpha(Y)T(ON)}) = K_{YZ}$$

  $$\Rightarrow\ T_{ON} = 1/\alpha_Y \cdot \ln[1/(1 - K_{YZ}/Y_{st})]$$

- Note that $T_{ON} \rightarrow \infty$ if $K_{YZ} \rightarrow Y_{st}$. Hence K_{YZ} should be clearly smaller than Y_{st}. In bacteria, K_{YZ} is typically 3 to 10 times lower that Y_{st} and delay T_{ON} is from few minutes to few hours

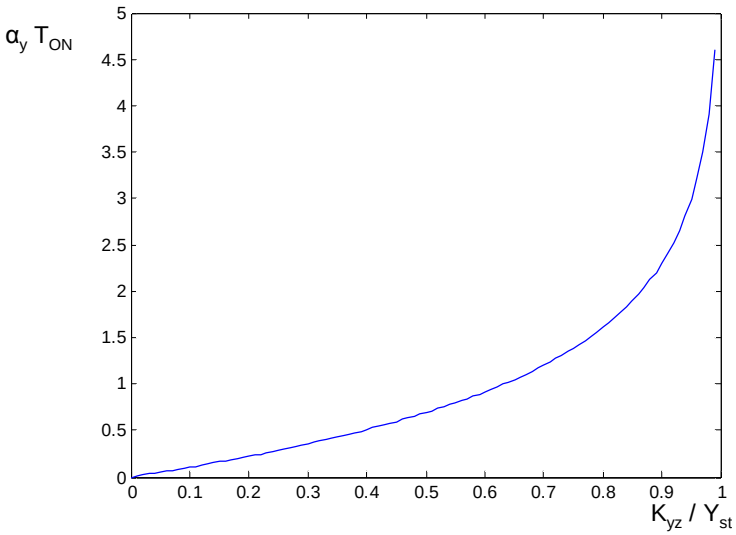


Fig 4.8a: Delay in Z production the C1-FFL following an ON step of inducer S_x as a function of the biochemical parameters of the transcription factor Y. The delay T_{ON}, made dimensionless by multiplying with the degradation/dilution rate of Y, α_y , is shown as a function of the ratio of the activation threshold K_{yz} and the maximal (steady-state) level of Y, denoted Y_{st}.

C1-FFL is a sign-sensitive delay element (cont.)

- So, there is a delay T_{ON} following an ON step of S_X, but
- No delay following an OFF step of S_X (see next slide, Fig 4.8 b)

- Hence: C1-FFL is a **sign-sensitive delay element**

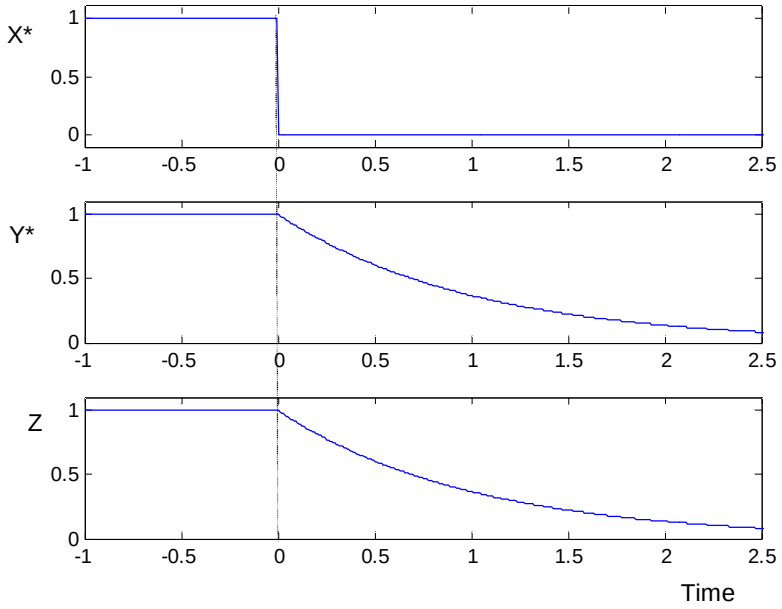


Fig 4.8 b: Dynamics of the C1-FFL following an OFF step of Sx at time t=0. All production and degradation rates are equal to 1.

4.6.4 Sign-sensitive delay can protect against brief input fluctuations

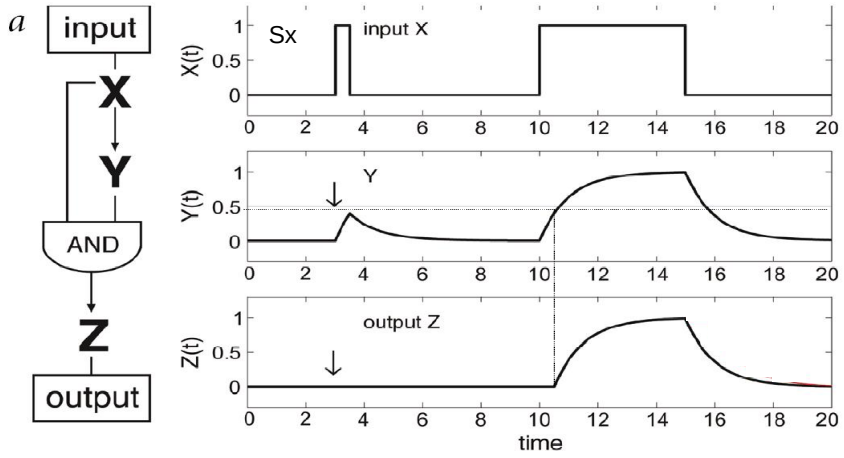


Fig 4.8c: The C1-FFL with AND-logic as a persistence detector.
A brief pulse of signal S_x does not give Y enough time to accumulate and cross its activation threshold for Z. Hence Z is not expressed. A persistent pulse yields Z expression at a delay. Z expression stops with no delay when Sx is removed.
Source: Shen-Orr 2002 Nature genetics.

4.6.5 Example: Sign-sensitive delay in the arabinose system of E. coli

Fig 4.9 Experimental dynamics of the C1-FFL in the arabinose system of E. coli. (See the figure on the next slide!)

a) The arabinose (ara) system encodes enzymes that utilize the sugar arabinose (araBAD), and transport it into the cell (araFGH). The system is activated upon glucose starvation by the activator X=CRP (signal Sx=cAMP, a molecule produced upon glucose starvation), and in the presence of Sy=arabinose by the activator Y=araC. The input function is an AND-gate.

As a control system with no FFL (simple regulation), the experiment used the lac operon, in which same activator X=CRP regulates the lactose operon, but X does not regulate Y1=LacI.

Dashed arrows: rapid, non-transcriptional positive and negative feedback loops in which the output gene products affect the signal (eg by transporting the sugar Sy into the cell and degrading it).

b) Dynamics of the promoter activity of the ara and the lac operons was monitored at high temporal resolution in growing cells by means of reporter-GFP fusions, in the presence of Sy. The experiments followed the dynamics after ON steps and OFF steps of Sx. Shown is GFP per cell normalized to its maximal level. A delay occurs in the FFL after on steps but not after OFF steps. Based on Mangan et al JMB 2003.

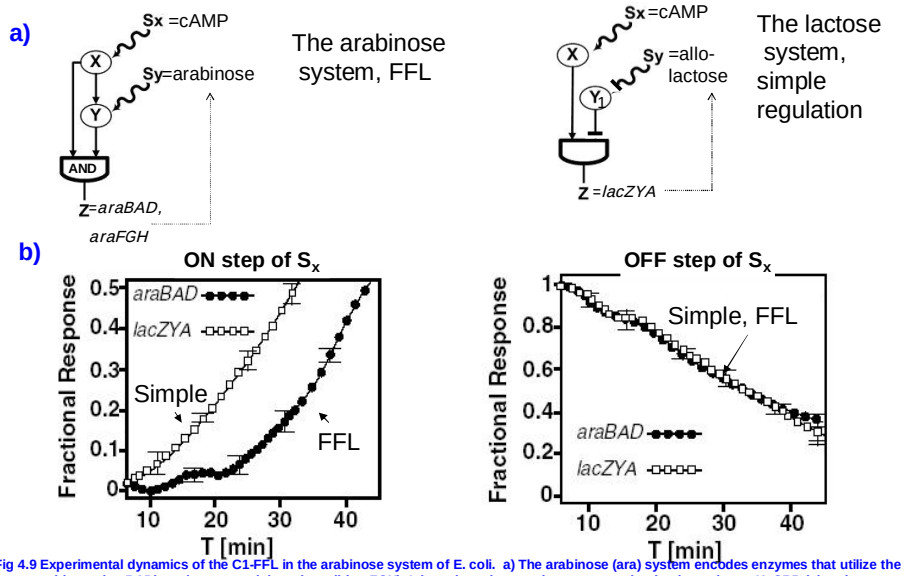


Fig 4.9 Experimental dynamics of the C1-FFL in the arabinose system of E. coli. a) The arabinose (ara) system encodes enzymes that utilize the sugar arabinose (araBAD), and transport it into the cell (araFGH). It is activated upon glucose starvation by the activator X=CRP (signal S_x=cAMP, a molecule produced upon glucose starvation), and in the presence of Sy=arabinose by the activator Y=araC. The input function is an AND-gate. As a control system with no FFL (simple regulation), the experiment used the lac operon, in which same activator X=CRP regulates the lactose operon, but X does not regulate Y_1=LacI. Dashed arrows: rapid, non-transcriptional positive and negative feedback loops in which the output gene products affect the signal (eg by transporting the sugar S_y into the cell and degrading it). b) Dynamics of the promoter activity of the ara and the lac operons was monitored at high temporal resolution in growing cells by means of reporter-GFP fusions, in the presence of S_y. The experiments followed the dynamics after ON steps and OFF steps of S_x. Shown is GFP per cell normalized to its maximal level. A delay occurs in the FFL after on steps but not after OFF steps. Based on Mangan et al JMB 2003.

4.6.6 The OR gate C1-FFL is a sign-sensitive delay for OFF steps of S_X

- C1-FFL with OR gate is a sign-sensitive delay element but with signs opposite to those of the AND version: a delay after OFF step, no delay after ON step (see Fig 4.10 EXTRA on the next slide)

- Fig 4.10 of the textbook (page 57):
  - FFL with OR gate that controls the production of proteins that self-assemble into a motor that rotates the flagella that allow *E. coli* to swim

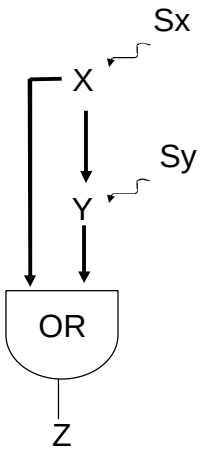


Fig 4.10 EXTRA. The C1-FFL with an OR input function: Transcription factor X activates the gene encoding transcription factor Y, and X or Y activate gene Z. The two input signals are Sx and Sy. An input-function integrates the effects of X and Y at the Z promoter (an OR-gate in this figure).

4.7 The incoherent type-1 FFL (I1-FFL)

4.7.1 Structure of the incoherent FFL

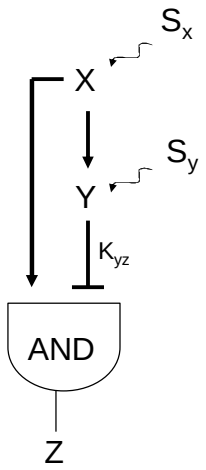


Fig 4.11 a: The incoherent type-1 FFL (I1-FFL) with an AND gate at the Z promoter. The inputs are the inducers S_x and S_y. The repression threshold of gene Z by repressor Y is K_{yz}.

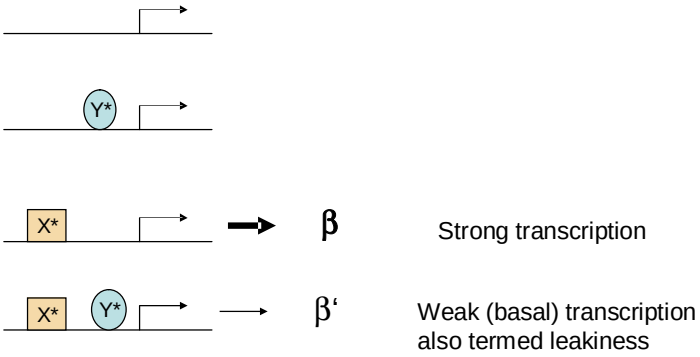


Fig 4.11 b: (I1-FFL cont.) The four binding states of a simple model for the promoter region of Z, regulated by activator X and repressor Y. Transcription occurs when the activator X* is bound, and to a much lesser extent when both activator and repressor Y* are bound.

4.7.2 Dynamics of the I1-FFL: A pulse generator

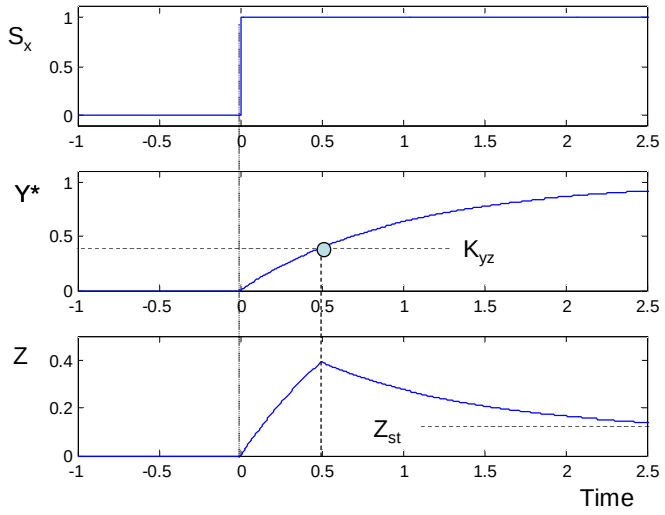


Fig 4.12 a: Dynamics of the I1-FFL with AND input function following an ON-step of S_x. The step occurs at t=0, and X rapidly transits to its active form X*. As a result, the repressor protein Y is produced, and represses Z production when it crosses the repression threshold K_{yz}. In this figure, all production and decay rates are equal to 1.

Dynamics of the I1-FFL: A pulse generator (cont.)

- Dynamics of Y: $dY/dt = \beta_Y - \alpha_Y Y^* \Rightarrow Y^*(t) = Y_{st}(1-e^{-\alpha(Y)t})$
- Dynamic equation of Z when $Y^* < K_{YZ}$: $dZ/dt = \beta_Z - \alpha_Z Z$
  $\Rightarrow Z(t) = Z_m(1 - e^{-\alpha(Z)t})$ where $Z_m = \beta_Z / \alpha_Z$
- When Y^* crosses K_{YZ}, the production rate of Z suddenly drops to low value β_Z' (= the basal production rate of Z); if no leakiness, $\beta_Z' = 0$.
- **Repression time** is defined as T_{rep}: $Y^*(T_{rep}) = K_{YZ}$
  $\Rightarrow T_{rep} = 1/\alpha_Y \ln[1/(1 - K_{YZ}/Y_{st})]$
- Level of Z decays to a lower steady-state $Z_{st} = \beta_Z'/\alpha_Z$
  $\Rightarrow Z(t) = Z_{st} + (Z_0 - Z_{st}) e^{-\alpha(Z)(1-T(rep))}$
  where Z_0 is the level at time $t=T_{rep}$: $Z_0 = Z_m(1 - e^{-\alpha(Z)T(rep)})$
- **Repression factor** is defined as $F = \beta_Z / \beta_Z' = Z_m/Z_{st}$

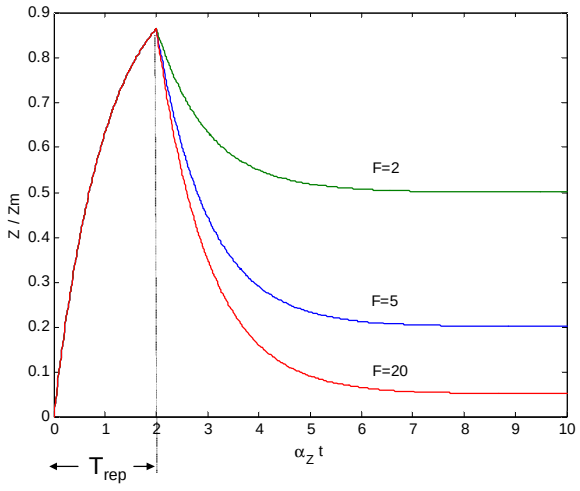


Fig 4.12 b: Expression dynamics of Z in an incoherent type-1 FFL with repression coefficients F = 2, 5, 20. The repression coefficient is the ratio of the maximal expression without active repressor to the steady-state expression with active repressor. T_{rep} is the time when repression begins, and is the moment of maximal Z concentration.

4.7.3 I1-FFL speeds the response time

- Fig 4.13 a (see next slide): response time of I1-FFL is shorter than that of a simple-regulation circuit with the same steady-state level of Z
- Response time $T_{1/2}$ for I1-FFL:
  Solving t from $Z_{st}/2 = Z_m(1-e^{-\alpha(Z)t})$ gives
  $T_{1/2} = 1/\alpha_Z \ln[2F/(2F-1)]$
  where F = repression factor Z_m / Z_{st}
- Fig 4.13 b: the larger is Z, the faster the response time becomes
- This response acceleration is sign-sensitive: no speed-up in the OFF direction

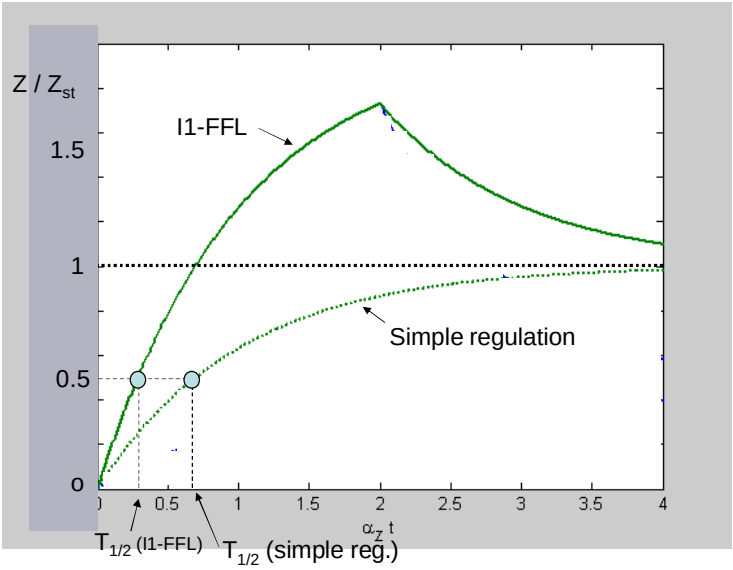
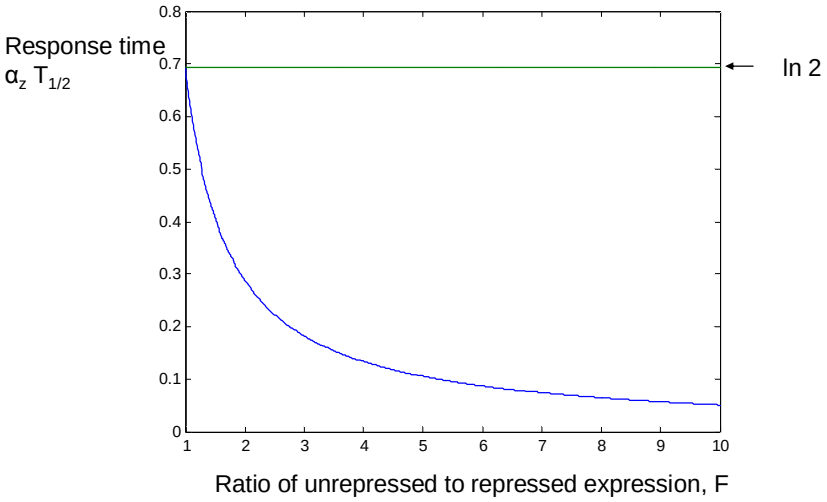


Figure 4.13 a: Response time of the I1-FFL is shorter than simple regulation that reaches same steady-state level. The normalized response time of simple regulation is ln 2 ~ 0.7. (Simple regulation - dashed line, I1-FFL - full line)

Figure 4.13 b: Response time of the I1-FFL as a function of the repression coefficient F. F is the ratio of unrepressed to repressed Z expression. Green horizontal line: normalized response time of simple regulation, $\alpha_z T_{1/2} = \ln 2$.



4.7.5 Experimental study of the dynamics of an I1-FFL

- Fig 4.14 (textbook page 64): Dynamics of the I1-FFL in the galactose system of *E. coli*.

4.7.6 Three ways to speed the response time (summary)

- Increased degradation rate
- Negative auto-regulation
- Incoherent FFL

4.8 Why are some FFL types rare?

- In I1-FFL, (the absence of) S_Y can turn on high expression $Z_{st} = \beta_Z/\alpha_Z$; see Fig 4.15 (next slide)
- In I4-FFL (and in I3-FFL as well), S_Y has no effect on Z_{st} which equals β_Z'/α_Z independently on whether or not S_Y is present (see Figs 4.16 and 4.17), that is, I4-FFL is not responsive to one of its two inputs (Table 4.3, page 67)
- The lack of responsiveness to one of the two inputs may be one of the reasons why I4-FFL and I3-FFL are selected less often (i.e., they are not network motifs) than I1-FFL

Fig 4.15: The effect of input signal S_y on the dynamics of the I1-FFL. When Sy is absent, Y is not active as a repressor, and the concentration of protein Z shows an increase to a high unrepressed steady-state (dashed line)

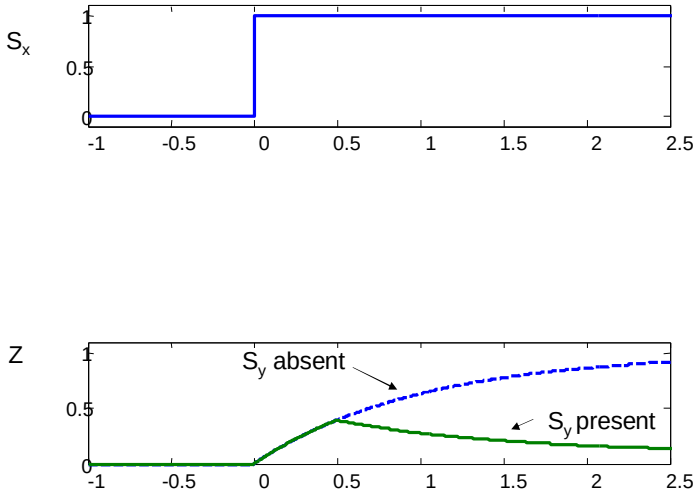
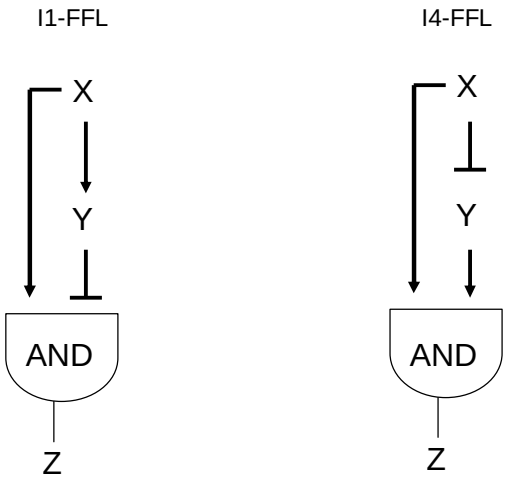


Fig 4.16: The incoherent types I1-FFL and I4-FFL



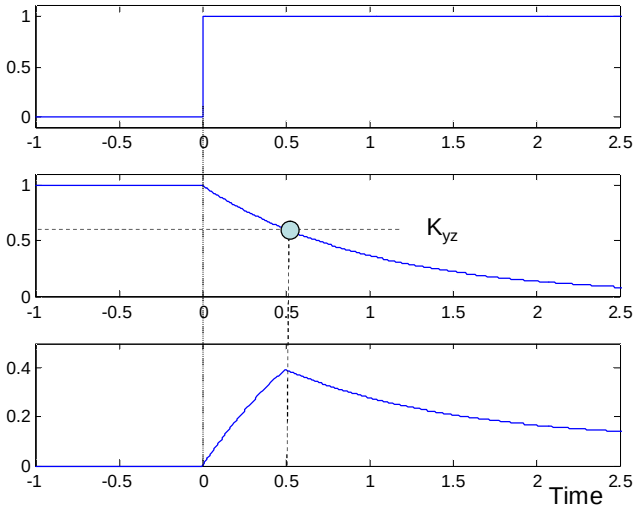


Fig 4.17: Dynamics of the I4-FFL following a step of Sx. In the presence of Sx, protein X is active and activates Z production but represses production of Y. When Y levels decay below the activation coefficient K_{yz}, the concentration of Z begins to drop. Production and decay rates are β_Z=1, α_z= α_y=1, F=10. The signal Sy is present throughout.

4.9 Convergent evolution of FFLs

- How did FFLs evolve?
- The most common form of evolution for genes is conservative evolution where two genes with similar function stem from a common ancestor gene
- It appears that in most cases the FFLs did not evolve in a similar way: homologous genes Z and Z' in two organisms are often both regulated by FFLs in response to environmental stimuli but their regulators X and Y, and X' and Y' are usually not homologous in such FFL pairs (see Fig 4.18 on the next slide)
- That is, evolution seems to have converged independently on the same regulation circuit in many cases (Conant and Wagner 2003, Babu et al 2004). Presumably, the FFL is rediscovered by evolution because it performs an important function in different organisms

Fig 4.18: On the evolution of the FFLs.

(a) The V-shaped pattern in which X and Y regulate Z is strongly selected because it allows regulation based on two inputs. The edge from X to Y (white arrow) must be selected based on the basis of an additional dynamical function (e.g. sign sensitive delay, acceleration, pulse generation).

(b) In many cases homologous genes Z and Z' in different organisms are regulated in a FFL in response to the same stimuli, but the two regulators X and Y in the FFL are not homologous to the regulators X' and Y'. Homology means sufficient similarity in the genes sequence to indicate that the genes have a common ancestor.