SOLUTIONS MANUAL FOR

AN INTRODUCTION TO SYSTEMS BIOLOGY
DESIGN PRINCIPLES OF BIOLOGICAL CIRCUITS

by

URI ALON

Chapman & Hall/CRC
Taylor & Francis Group

SOLUTIONS MANUAL FOR

AN INTRODUCTION TO SYSTEMS BIOLOGY
DESIGN PRINCIPLES OF BIOLOGICAL CIRCUITS

by

URI ALON

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Table of Contents

This page intentionally left blank

Exercises, chapter 2

2.1 *A change in production rate.* A gene Y with simple regulation is produced at a constant rate β_1. The production rate suddenly shifts to a different rate β_2 .
(a)Calculate and plot Y(t). (b)What is the response time (time to reach halfway between the steady-states)?

Solution:

(a): Let's mark the time when the shift occurs as t=0. Before the shift, Y reaches steady state at a level $Y(t=0)=Y_{st}= \beta_1 / \alpha$. After the shift,

(2.1.1) $dY/dt= \beta_2 - \alpha Y$.

The solution of such an equation is generally $Y = C_1 + C_2 \exp(-\alpha t)$, where the constants C_1 and C_2 need to be determined so that $Y(t=0)= \beta_1 / \alpha$, and Y at long times reaches its new steady state β_2 / α. This yields the following sum of an exponential and a constant

(2.1.2) $Y(t) = \beta_1 / \alpha + (\beta_2 / \alpha - \beta_1 / \alpha)(1-e^{-\alpha t}) = \beta_2 / \alpha + (\beta_1 / \alpha - \beta_2 / \alpha) e^{-\alpha t}$

Take the derivative with respect to time, dY/dt, and verify that Eq. 2.1.1 is fulfilled.
(b) The response time, which is the time to reach half way between the two steady states, is $\log(2)/\alpha$.

2.2 *mRNA dynamics.* In the main text, we considered the activation of transcription of a gene (mRNA production), and used a dynamical equation to describe the changes in the concentration of the gene product, the concentration of protein Y: $dY/dt= \beta - \alpha Y$, in which β describes the rate of protein production. In reality, mRNA needs to be translated to form the protein, and mRNA itself is also degraded by specific enzymes.
(a) Derive dynamical equations for the rate of change of mRNA and the rate of change of the protein product, assuming that mRNA is produced at rate β_m and degraded at rate α_m, and that each mRNA produces on average p protein molecules per second. The protein product is degraded/diluted at rate α.
(b) Note that mRNA is often degraded much faster than the protein product $\alpha_m >> \alpha$ (in bacteria, mRNA lifetime is usually on the order of minutes (Bernstein et al., 2002) whereas most proteins are stable for hours). Can this be used to form a quasi-steady-state assumption that mRNA levels are at steady-state with respect to slower processes? What is the effective protein production rate β in terms of β_m, α_m and p?

What would be the response time if the mRNA lifetime were much longer than the protein lifetime?

Solution:

(a) The dynamic equation for the concentration of mRNA of gene Y, Y_m, is

(2.2.1) $dY_m / dt = \beta_m - \alpha_m Y_m$.

The dynamical equation for the protein product is due to production of p copies per mRNA and degradation/dilution at rate α:

(2.2.2) $dY / dt = p Y_m - \alpha Y$

(b) In the typical case that mRNA degradation is faster than the degradation/dilution of the protein product, we can assume that Y_m reaches steady-state quickly in comparison to the protein levels. The reason is that the typical time for the mRNA to reach steady state is the response time $\log(2)/ \alpha_m$, which is much shorter than the protein response time $\log(2)/ \alpha$ because $\alpha_m >> \alpha$. The steady-state mRNA level is found by setting $dY_m / dt=0$ in Eq. 2.2.1 , yielding

(2.2.3) $Y_{m, st} = \beta_m / \alpha_m$

Using this for Y_m in Eq 2.2.2 yields the following equation for the protein production rate

(2.2.4) $dY / dt = p \beta_m / \alpha_m - \alpha Y$

In other words, the effective protein production rate, which is the first term on the right hand side of the equation, is equal to the steady state mRNA level times the number of proteins translated from each mRNA

(2.2.5) $\beta = p \beta_m / \alpha_m$

In cases where $\alpha_m << \alpha$, that is when mRNA is much more stable than the protein, the response time is governed by the slower process, accumulation of mRNA. For each level of mRNA, $Y_m(t)$, protein level 'instantly' reaches its momentary steady state $pY_m(t)/\alpha$. In such a case the response time is $\log(2)/\alpha_m$ according to the same reasoning as in the text for the protein response time. For advanced students, a full solution of the dynamics is given at the end (not necessary for the solution required in this exercise).

2.3 *Time-dependent production and decay.* A gene Y with simple regulation has a time-dependent production rate $\beta(t)$ and a time-dependent degradation rate $\alpha(t)$. Solve for its concentration as a function of time.

Solution:

Verify by taking the time derivative that the following is correct:

(2.3.1) $$Y(t) = e^{-\int \alpha(t')dt'}[Y(0) + \int \beta(t')e^{\int \alpha(t'')dt''}dt']$$

where all integrals are between 0 and t.

For example, for the case of problem 2.1, $\alpha(t)$ is constant over time so that $exp(-\int \alpha(t')\,dt') = \exp(-\alpha t)$, and $\beta(t')$ is constant after t=0 and equal to β_2 so that:

$$\int \beta(t')e^{\int \alpha(t'')dt''}dt' = \int \beta_2 e^{\alpha t'}dt' = \frac{\beta_2}{\alpha}(e^{\alpha t} - 1) ,$$

and we obtain the desired result $Y(t)=\beta_2 / \alpha + (\beta_1 / \alpha - \beta_2 / \alpha) \exp(-\alpha t)$.

2.4 *Cascades.* Consider a cascade of three activators, X→Y→Z. Protein X is initially present in the cell in its inactive from. The input signal of X, S_x, appears at time t=0. As a result, X rapidly becomes active and binds the promoter of gene Y, so that protein Y starts to be produced at rate β. When Y levels exceed a threshold K_y, gene Z begins to be transcribed. What is the concentration of gene product Z as a function of time? What is its response-time with respect to addition of S_x? What about a cascade of three repressors? Compare your solution to the experiments shown in Fig 2.7.

Solution:

We will assume all proteins have the same dilution/degredation rate α. After induction, Y is produced at rate

β_y and degraded/diluted at rate α:

(2.4.1) $$dY/dt=\beta_y-\alpha Y$$

yielding the familiar exponential approach to steady-state:

$Y(t)=\beta_y/\alpha\,(1-\exp(-\alpha t))$

Assuming a step function for the activation of gene Z by Y (logic input function), transcription of gene Z starts at time τ_{yz} when $Y(\tau_{yz})=K_y$:

(2.4.2) $$Y(\tau_{yz})=\beta_y/\alpha\,(1-\exp(-\alpha\tau_{yz}))=K_y \Longrightarrow \tau_{yz}=1/\alpha \log(Y_{st}/(Y_{st}-K_y))$$

where $Y_{st}=\beta_y/\alpha$. Just for extra clarity, let's consider the limits of (2.4.2) to see if this makes sense. When $K_y<<Y_{st}$, $Y_{st}-K_y \longrightarrow Y_{st}$ and $\tau_{yz} \longrightarrow 0$. In this case the threshold for Z activation is low, and Y levels cross it very fast. Conversely, if the activation

threshold K_y is very high, approaching Y_{st}, Z is never activated because $Y_{st}-K_y \longrightarrow 0$ and $\tau_{yz} \longrightarrow \infty$.

Production of Z starts after time $t=\tau_{yz}$ at a constant rate of β_z:

(2.4.3) $dZ/dt = 0 \quad (t<\tau_{yz})$

$\beta_z - \alpha Z \quad (t>\tau_{yz})$

Solving this we get:

$Z(t) = 0 \quad (t<\tau_{yz})$

$\beta_z/\alpha\,(1-\exp(-\alpha(t-\tau_{yz}))) \quad (t>\tau_{yz})$

Solving for the response time, the time to reach half of the steady state of Z:

(2.4.4) $\beta_z/\alpha\,(1-\exp(-\alpha(t_{1/2}-\tau_{yz}))) = 1/2\,\beta_z/\alpha \Longrightarrow$

$t_{1/2}=\tau_{yz}+\log(2)/\alpha$

Hence, there is an extra delay of τ_{yz} in the response time of gene Z relative to simple regulation with no cascade.

If Z activates a third gene W when it crosses a threshold K_z, this will occur at a time τ_{zw} found from:

(2.4.5) $\beta_z/\alpha\,(1-\exp(-\alpha(\tau_{zw}-\tau_{yz})))= Z_{st}\,(1-\exp(-\alpha(\tau_{zw}-\tau_{yz}))) = K_z$

solving for τ_{zw} we obtain:

(2.4.6) $t_{zw}=t_{yz}+1/\alpha\,\log(Z_{st}/(Z_{st}-K_z))$

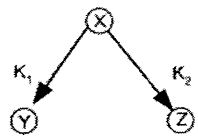
We can generalize this result: each step in a cascade, where a gene X activates a downstream gene after crossing a threshold K_x adds a delay of :

(2.4.7) $\tau_{delay}=1/\alpha\,\log(X_{st}/(X_{st}-K_x))$

In the special case in which the activation threshold is half the steady-state level (this can be shown to be in some cases an optimal value), the delay is $\tau_{delay}=\log(2)/\alpha$. In summary, since $1/\alpha$ is often on the scale of a cell generation, a transcriptional cascade can be a slow process.

2.5 *Fan-out:* Transcription factor X regulates two genes Y_1 and Y_2. Draw the resulting network, termed a fan-out with two target genes. The activation thresholds

for these genes are K_1 and K_2. The activator X begins to be produced at time $t=0$ at rate β, and is degraded/diluted at rate α, and its signal S_x is present throughout. What are the times at which Y_1 and Y_2 reach halfway to their maximal expression? Design a fan-out with three target genes in which the genes are activated with equal temporal spacing.



Based on problem 2.4:

(2.5.1) $\tau_1=1/\alpha \log(X_{st}/(X_{st}-K_1))$, $\tau_2=1/\alpha \log(X_{st}/(X_{st}-K_2))$

After the corresponding delays in gene activation, denoted τ_1 and τ_2, production of Y_1 and Y_2 starts at a constant rate reaching half the steady state after $\log(2)/\alpha$. The time to reach half maximum is therefore: $t_{1/2}=\tau_i+\log(2)/\alpha$ (i=1,2), where i=1,2 for Y_1 and Y_2 respectively.

For three target genes, we require $\tau_2=1/2\ (\tau_1+\tau_3)$. This amounts to the following requirements on the thresholds,

(2.5.2) $1/\alpha \log(X_{st}/(X_{st}-K_2))=1/2\ (1/\alpha \log(X_{st}/(X_{st}-K_1))+ 1/\alpha \log(X_{st}/(X_{st}-K_3)))$

$$\Longrightarrow \quad X_{st}-K_2=\sqrt{(X_{st}-K_1)(X_{st}-K_3)}$$

2.6 Pulse of activation: Consider the cascade of exercise 2.4. The input signal S_x appears at time t=0 for a pulse of duration D, and then vanishes.

(a) What is the concentration Y(t)?

(b) What is the minimal pulse duration needed for activation of gene Z?

(c) Plot the maximal level reached by the gene product Z as a function of the pulse duration D.

Solution:

a) Protein X*, the active conformation of protein X bound to its inducer, binds the promoter of Y commencing its production according to (2.4.1) yielding the familiar exponential approach to steady-state:

(2.6.1) $Y(t)=\beta_y/\alpha\,(1-\exp(-\alpha t))$

b) Protein Z will be activated when Y levels cross the threshold K_{YZ}:

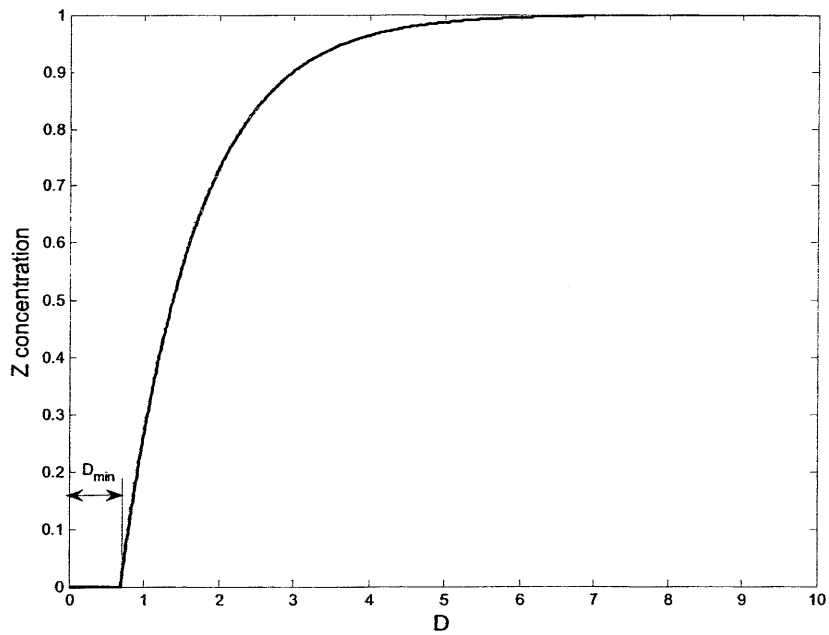
(2.6.2) $Y(t)=\beta_y/\alpha\,(1-\exp(-\alpha D_{min}))= Y_{st}\,(1-\exp(-\alpha D_{min}))=K_{YZ}$

Solving for D:

(2.6.3) $D_{min}=1/\alpha\,\log\,(1/(1-K_{YZ}/Y_{st}))$

Let's consider the limits of (2.6.3). If K_{YZ} is very small D approaches 0, and a very short duration of the input S_x suffices to activate Z. When K_{YZ} approaches Y_{st} D approaches infinity.

c)



The schematic plot uses $K_{YZ}=0.5*Y_{st}$, and $\alpha=1$. After the delay of D_{min} Z concentration follows the familiar exponential rise to steady state. In summary, the 3-gene cascade results in a delay in the activation of the Z gene following an input to the X gene.

Appendix

Full solution of the dynamics of problem 2.2

The full solution of this problem can be obtained by inserting the time-dependent solution of equations 2.2.1 into equation 2.2.2 to obtain:

(2.A.1) $dY/dt=p\ Y_m-\alpha Y=p\beta_m/\alpha_m\ (1-\exp(-\alpha_m t))-\alpha Y$

Using the method of problem 2.3 (below) the full solution for Y(t) is:

(2.A.2) $$Y(t)=\frac{p\beta_m}{\alpha_m(\alpha-\alpha_m)}(1-\exp(-\alpha_m t))-\frac{p\beta_m}{\alpha(\alpha-\alpha_m)}(1-\exp(-\alpha t))$$

In the limit where mRNA degradation is much higher than protein degradation and for $1/\alpha_m<<t$, we can approximate $(\alpha-\alpha_m)\sim-\alpha_m$ and $\exp(-\alpha_m t)\sim 0$. Equation (2.A.2) then reduces to:

(2.A.3) $$Y(t)=\frac{p\beta_m}{\alpha\,\alpha_m}(1-\exp(-\alpha t))-\frac{p\beta_m}{\alpha_m^2}\sim\frac{p\beta_m}{\alpha\,\alpha_m}(1-\exp(-\alpha t))$$

In the limit where mRNA degradation is much smaller than protein degradation and for $1/\alpha<<t$, we can approximate $(\alpha-\alpha_m)\sim\alpha$ and $\exp(-\alpha t)\sim 0$. Equation (2.A.2) then reduces to:

(2.A.4) $$Y(t)=\frac{p\beta_m}{\alpha_m\alpha}(1-\exp(-\alpha_m t))-\frac{p\beta_m}{\alpha^2}\sim\frac{p\beta_m}{\alpha_m\alpha}(1-\exp(-\alpha_m t))$$

The steady state solution for Y is the same in both of these limiting cases : $Y_{st}=\ p\ \beta_m\ /\ \alpha\alpha_m$, but the response time is governed by the slower process of the two degradation processes – if mRNA degradation is much higher than protein degradation, the response time is governed by protein degradation: $t_{1/2}=\log(2)/\alpha$. If mRNA degradation is much smaller than protein degradation, the response time is governed by mRNA degradation $t_{1/2}=\log(2)/\alpha_m$.

Problem 2.3 – integration factors

A general solution of a first order linear differential equation of the following form:

(2.A.5) $dY/dt+\alpha(t)Y=\beta(t)$

can be obtained by first multiplying both sides by an integration factor $\mu(t)$ defined as:

(2.A.6) $\mu(t)=exp(\int\alpha(t')dt')$

This gives us:

(2.A.7) $\mu(dY/dt)+\alpha\mu Y=\beta\mu$

The left side can be replaced by:

(2.A.8) $d(\mu Y)/dt=\beta\mu$

Integrating this equation:

(2.A.9) $\mu Y = \int \beta \mu dt$

which leads to

(2.A.10) $Y(t) = 1/\mu \, [Y(0) + \int \beta \mu dt \,]$

Finally, inserting μ from 2.A.6 we get equation 2.3.1

(2.A.11) $Y(t) = exp(-\int \alpha(t')dt') \, [Y(0) + \int \beta(t') \, exp(\int \alpha(t'')dt'') \, dt' \,]$

Exercises, chapter 3

3.1. *Auto-repression with Hill input function.* What is the response time for a repressor that cooperatively represses its own promoter (described by a Hill-function with Hill coefficient *n*):

(3.1.1) $dX / dt = \beta / (1+(X / K)^n) - \alpha X$

How much faster is the response than in non-auto-regulated circuits? Use the approximation of strong auto-repression, that is $(X / K)^n >> 1$.

Solution:

In the limit of strong auto-repression, we can neglect the '1' in the denominator of the dynamic equation as soon as $(X / K)^n >> 1$, and we have[1]

(3.1.2) $dX/dt = \beta K^n / X^n - \alpha X$

To solve this equation, multiply both sides by X^n and switch to the new variable $u=X^{n+1}$. Note that $d u / dt = (n+1) X^n dX /dt$. The equation now reads

(3.1.3) $d u /dt =(n+1) \beta K^n - (n+1) \alpha u.$

The solution of this linear equation is simple exponential convergence to steady-state, the same as in chapter 2 :

(3.1.4) $u = u_{st} (1-\exp(- (n+1) \alpha t)).$

Switching back to the original variable X, we have

(3.1.5) $X=X_{st} (1-\exp(- (n+1) \alpha t))^{1/(n+1)}$

The response time is found by $X(T_{1/2}) = X_{st} / 2$. This yields

(3.1.6) $T_{1/2} = 1/[(n+1) \alpha] \log(2^{n+1}/[2^{n+1}-1]).$

The response time becomes shorter the larger n. For n=1, 2, 3, the ratio of $T_{1/2}$ to the response time of simply regulated genes ($\log(2)/ \alpha$) is about $T_{1/2}^{(n.a.r)}/ T_{1/2}^{(simple)} = 0.2$, 0.06 and 0.02 respectively. See Fig 3.5 for the dynamics of a strongly negatively-autoregulated gene. The sharper the negative auto-regulation (higher n), the more the

[1] When is this approximation valid? Note that the steady state is, according to Eq 3.1.2, $X_{st}= K (\beta/ \alpha K)^{1/(n+1)}$. Thus when the unrepressed steady-state is much higher than the repression coefficient, that is when $\beta/ \alpha >> K$, we have $X_{st}>>K$. This means that to describe the dynamics that occur when X exceeds K and begins to approach its steady state, we can neglect the '1' in the denominator of the input function.

system approaches the sharp logic-function (Boolean) limit discussed in this chapter, and the faster it responds.

3.2 *Parameter sensitivity*: Analyze the robustness of the steady-state with respect to cell-cell variations in the production rate β for the system of problem 3.1. Calculate the parameter-sensitivity coefficient (Savageau, 1976) (Heinrich and Schuster, 1996) of the steady-state concentration with respect to β. The **parameter-sensitivity coefficient** of property A with respect to parameter B, denoted S(A,B), is defined as the relative change in A for a given, small relative change in B, that is S

(3.2.1) $S(A,B)=(\Delta A / A) / (\Delta B / B) = (B / A)\, dA/dB$

Solution:

The steady state level is found from Eq 3.1.2 using dX/dt=0, yielding

(3.2.2) $X_{st} = K (\beta/\alpha K)^{1/(n+1)}$.

The parameter-sensitivity which describes relative changes in steady state due to changes in production rate is

(3.2.3) $S(X_{st}, \beta) =(\beta / X_{st})\, d X_{st} / d \beta = 1/(n+1)$.

Thus, sensitivity decreases with Hill coefficient n. The higher n, the weaker the dependence of the steady-state on β and α. In other words, robustness to variations in production rates increases with the Hill coefficient.

For Hill coefficient n=4, for example, $S(X_{st}, \beta) =1/5$, which means that a 10% change in β yields only a 2% change in X_{st}. In the limit of very high n, the steady state does not depend at all on production or degradation rates, $X_{st}=K$. This is the steady-state solution found in the main text for the logic input-function. Simple regulation is equivalent to n=0, so that S(Xst, β)=1. This means that a small change of x% in production leads to the same change of x% in steady-state.

3.3 Autoregulated cascade: Gene X encodes a repressor that represses gene Y, which also encodes a repressor. Both X and Y negatively regulate their own promoters.

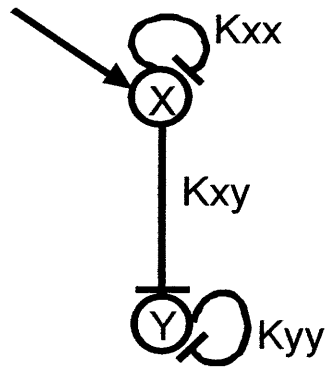
(a) At time t=0, X begins to be produced at rate β, starting from an initial concentration of X=0. What are the dynamics of X and Y? What are the response times of X and Y? Assume logic input functions, with repression thresholds K_{xx}, K_{xy}

for the action of X on its own promoter and on the Y promoter and K_{yy} for the action of Y on its own promoter.

(b) At time t=0, production of X stops after a long period of production, and X concentration decays from an initial steady-state level. What are the dynamics of X and Y? What are the response times of X and Y?

Solution:

The following diagram describes the circuit. We assume X has an additional activator that can turn on its production (represented by the incoming edge at X).



a) Assuming logic input function in all promoters:

(3.3.1) $dX/dt=\beta_x\,\theta(X<K_{xx})-\alpha X$

$dY/dt=\beta_y\,\theta(X<K_{xy})\,\theta(Y<K_{yy})-\alpha Y$

X concentration initially follows the familiar exponential rise (as long as $X<K_{xx}$):

(3.3.2) $X(t)=\beta_x/\alpha\,(1-\exp(-\alpha t))$

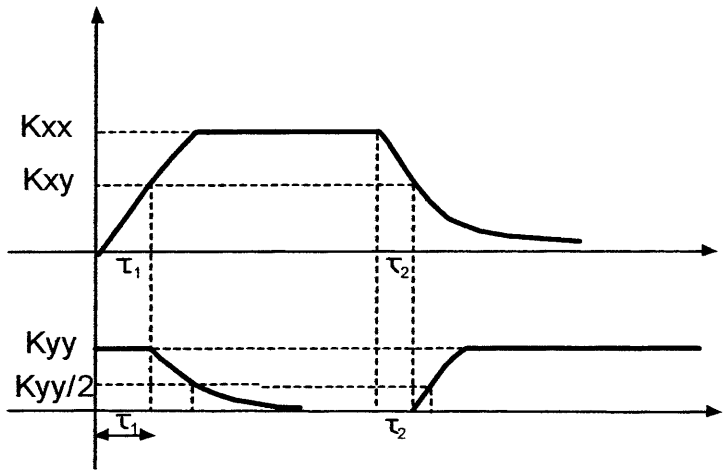
When $X(\tau_1)=K_{xy}$ Y production stops and it exponentially decays from its initial steady state value of K_{yy} toward zero. The delay is:

(3.3.3) $X(\tau_1)=\beta_x/\alpha\,(1-\exp(-\alpha\tau_1))=K_{xy} \Longrightarrow \tau_1=1/\alpha\,\log(\beta_x/(\beta_x-\alpha K_{xy}))$

In the limit of strong auto-repression, $\tau_1=K_{xy}/\beta_x$ The time for Y(t) to reach half its steady-state value ($K_{yy}/2$) is:

$t_{1/2}=\tau_1+\log(2)/\alpha$

Note that Y's negative auto-regulation has no effect here (see problem 3.5 below).



b) If X production stops (For example if its activator becomes in-active) its level will exponentially decay from its steady state level K_{xx} towards zero. At a delay τ_2 it will cross K_{xy}:

(3.3.4) $X(\tau_2)=K_{xx}\exp(-\alpha\tau_2)=K_{xy} \implies \tau_2=1/\alpha \log(K_{xx}/K_{xy})$

After τ_2 Y is produced at rate β_2 and will reach half of its steady state level K_{yy} after:

$t_{1/2} = \tau_2+K_{yy}/2\beta_y$

In the figure above representing the dynamics we assume, as in the text the limit of strong auto-repression, $K_{xx}<<\beta_x/\alpha$, and that $K_{yy}<<\beta_y/\alpha$.

3.4. *Positive feedback.* What is the effect of *positive* auto-regulation on the response time? Use as a model the following linear equation:

$dX / dt = \beta + \beta_1 X - \alpha X$

Explain each term and solve for the response time. When might such a design be biologically useful?

Solution: The basal production rate is β, the positive effect of X on its own production (positive auto-regulation) is described in this model by the linear term $\beta_1 X$, and degradation/dilution is represented as usual by $-\alpha X$.

Let's group the terms that multiply X in this linear model:

(3.4.1) $dX / dt = \beta - (\alpha - \beta_1) X$

We see that the degradation/dilution rate is effectively reduced by positive auto-regulation, to an effective rate $\alpha' = \alpha - \beta_1$. Assuming that the auto-regulation is not too strong, that is that $\beta_1 < \alpha$, the term multiplying X is negative and we get an approach to a stable steady-state:

(3.4.2) $X(t)= X_{st} (1- \exp(-\alpha ' t))$.

Where $\alpha' = \alpha - \beta_1$

The response time is defined as the time to reach half of the steady state:

$T_{1/2}$ (positive-autoregulation) $=\log(2)/\alpha'$

The response time is *longer* than that for simple regulation due to the reduced effective degradation/dilution rate:

$T_{1/2}$(positive-auto-regulation)$=\log(2)/(\alpha - \beta_1) > \log(2)/\alpha = T_{1/2}$ (simple regulation).

Thus positive auto-regulation has an effect that is opposite to that of negative auto-regulation. The former slows response time, whereas the latter speeds response times.

Note that strong auto-regulation, in which $\beta_1 > \alpha$, can lead to instability and unchecked growth of X in the model. In real systems, this instability will be limited by other factors (such as saturation of the input function), locking X in an ON state of high expression even after its activating input β vanishes.

Hence, strong positive feedback creates a bi-stable system, in which X is either at a low or at a high fixed point. This is useful for commitment-type biological decisions, such as those made in development. Positive feedback characterizes developmental systems that make a switch that is either OFF or is locked ON (e.g., a cell commits to become a muscle cell rather than, say , a blood cell, by means of positive feedback loops on key transcription factors).

A different biological example is found in some regulatory systems that govern the transcription of protein parts of multi-protein structures that are assembled slowly. An example is the bacterial flagellum described in Chapter 6 that can take two cell generations to be completed. Such slow processes can benefit from weak positive auto-regulation to slow down responses and prolong delays (Kalir et al., 2005).

3.5 *Turning off auto-regulation.* What is the dynamics of a negatively-auto-regulated gene once its maximal promoter activity is suddenly reduced from β_1 to $\beta_2=0$? What is the response time, and how does it compare to simple regulation?

Solution:

A negatively auto-regulated gene, the promoter activity of which has reduced to $\beta=0$, exponentially decays to zero with a response time of $\log(2)/\alpha$. The negative auto-regulation has no effect in this case. Assuming an additional activator is present which activates transcription at a constant rate β when it is present (ON step), negative auto-

regulation thus has an asymmetric accelerating effect, decreasing the response time for an ON signal, while not affecting the response time for an OFF signal.

3.6 *Two-node positive feedback for decision making:* During development from an egg to an embryo, cells need to make irreversible decisions to express the genes appropriate to their designated tissue types, and repress other genes. One common mechanism is positive transcriptional feedback between two genes. There are two types of positive feedback made of two transcription factors. The first type is of two positive interactions X→Y and Y→X. The second type has two negative interactions X--|Y and Y--|X. What are the stable steady states in each type of feedback? Which type of feedback would be useful in situations where genes regulated by both X and Y belong to the same tissue? Which would be useful when genes regulated by X belong to different tissues than the genes regulated by Y?

Solution:

Positive feedback with two positive interactions has two stable steady-states: X and Y both ON, and the reverse, X and Y both OFF. This is useful when genes regulated by X and genes regulated by Y belong to the same tissue. Positive feedback with two negative interactions has two stable steady states: X ON and Y OFF, and the reverse, X OFF and Y ON. That is either Y or X is ON. This is useful when genes regulated by X belong to different tissues or cell fates than the genes regulated by Y.

Excercises, chapter 4

4.1 *The second input.* What is the affect of steps of S_y on the expression dynamics of Z in the C1-FFL with AND logic? Are there delays in Z expression for ON or OFF steps of S_y? What is the response time of Z for such steps? Assume that S_x is present throughout.

Solution:

Assuming X and Y are at their steady-state level, a step of S_y will cause an immediate production of Z, which will follow the exponential dynamics towards steady-state. An OFF step of S_y will immediately stop Z production, causing it to decay exponentially towards its basal level. Hence in the C1-FFL with AND logic input function at the Z promoter, steps of S_y do not change the response times of Z, relative to a simple $Y \longrightarrow Z$ regulation.

4.2 OR-gate logic. Analyze the C1-FFL with OR-logic at the Z promoter. Are there delays following ON or OFF steps of S_x? What could be the biological use of such a design?

Solution:

After an ON step of S_x, X becomes active X*. On a rapid timescale it binds the Z promoter. Since Z is regulated by OR-gate logic, X* alone can activate transcription without need of Y. Therefore there are no delays following an ON step of S_x.

After an OFF step of S_x, X* rapidly becomes inactive X. However, protein Y is still present in the cell, and if S_y is present, Y is active Y*. Since the Z input function is an OR gate, Y* can continue to activate transcription of Z even in the absence of X*. Therefore, Z production persists, until Y degrades/dilutes below its activation threshold for Z. The dynamics of Y are given by

$dY / dt = -\alpha Y$, (there is no production term because X is inactive following removal of S_x) so that $Y=Y_m \exp(-\alpha t)$, where Y_m is the level of Y at time t=0.

The OFF delay is given by the time it takes Y to reach its activation threshold for Z, K_y: solving for this time

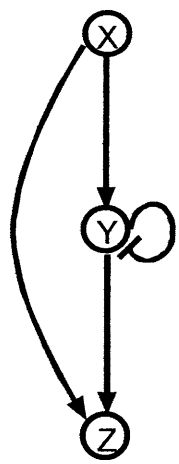
$Y(T_{off})= Y_m \exp(-\alpha T_{off})=K_y$, yields

$T_{off} = 1/\alpha \log (Y_m/K_y)$

In summary, the OR-gate C1-FFL shows sign-sensitive delays. It has a delay following OFF but not ON steps of S_x. The delay depends on the presence of Sy. This behavior is opposite to that of the C1-FFL with an AND gate, which shows delay upon ON but not OFF steps.

The OR-gate C1-FFL could be useful in systems which need to be protected from sudden loss of activity of their master regulator X. The OR-gate FFL can provide continued production during brief fluctuations in which X activity is lost. This protection works for OFF pulses shorter than T_{off}. Note that T_{off} can be tuned by evolutionary selection by adjusting the biochemical parameters of protein Y such as its expression level Y_m and its activation threshold K_y.

(4.3) *A decoration on the FFL.* The regulator Y in C1-FFLs in transcription networks is often negatively auto-regulated. How does this affect the dynamics of the circuit, assuming that it has an AND input function at the Z promoter? How does it affect the delay times? The Y regulator in an OR-gate C1-FFL is often positively auto-regulated. How does this affect the dynamics of the circuit? How does it affect the delay times?



Solution:

The negative auto-regulation on Y speeds its response time. Hence it shortens the time needed for Y to cross K_{yz}, the activation threshold for Z. Denoting by T_{on} the delay in Z activation following a step of S_x, and assuming strong auto-repression in the negative auto-regulation loop, the delay is:

$T_{on}=K_{yz}/\beta$ (see Eq 3.4.7), which is shorter than the delay in an FFL without negative auto-regulation. . Following an OFF step of S_x the negative auto-regulation has no effect (see Exercise 3.5) .

For the C1-FFL with an OR input function at the Z promoter, an ON step of S_x causes an immediate rise of Z (the auto-regulation of Y has no effect here because one active regulator input to Z is enough). The positive auto-regulation of Y has an effect upon an OFF step of S_x. Recall from exercise 3.4, that a positively auto-regulated gene has, in a linear model, the following dynamics:

(4.3.1) $dY / dt = \beta_x+\beta_1 Y -\alpha Y= \beta_x-(\alpha-\beta_1)Y$

where $\beta_1 Y$ is the term representing the positive auto-regulation and β_x is the effect of X. When S_x goes to zero, $\beta_x=0$ and the solution for Y's dynamics is a decay from an initial value Y_{st}:

(4.3.2) $Y= Y_{st} \exp(-(\alpha-\beta_1)t)=Y_{st} \exp(-\alpha' t)$

thus regulator Y's levels will exponentially decay with a rate $\alpha' = \alpha- \beta_1$, smaller than the rate α for a gene without positive auto-regulation. Hence, the delay for turning off gene Z will be longer due to positive auto-regulation. Gene Z production will stop when Y decreases below its activation threshold:

(4.3.3) $Y(T_{off})=K_{yz} \Longrightarrow Y_{st} \exp(-\alpha' T_{off})=K_{yz} \Longrightarrow T_{off} =1/\alpha' \log (K_{yz}/Y_{st})$

The delay in the turning off of Z is therefore α/α' longer than the delay when Y is not auto-regulated.

(4.4) *The diamond.* The four-node diamond pattern occurs when X regulates Y and Z, and both Y and Z regulate gene W.

(a) How does the mean number of diamonds scale with network size in random ER networks?

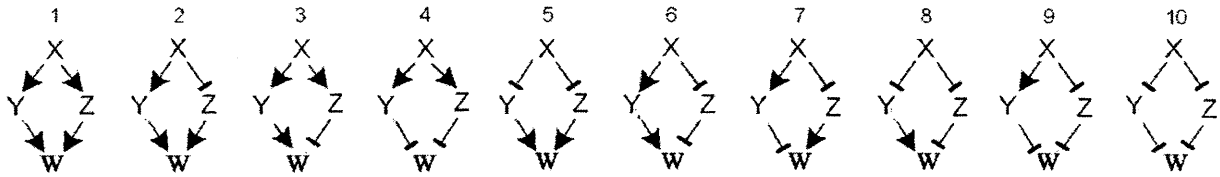
(b) What are the distinct types of sign combinations of the diamond (where each edge is either activation + or repression -)? How many of these are coherent? (Answer: 10 types, of which 6 are coherent).

(c) Consider a diamond with four activation edges. Assign activation thresholds to all edges. Analyze the dynamics of W following a step of Sx, for both AND and OR logic at the W promoter. Are there sign-sensitive delays?

Solution:

(a) the diamond has n=4 nodes and g=4 edges. The number of diamonds therefore scales in ER networks as $N^{n-g} = N^0$. Hence the number of diamonds does not depend on the ER network size.

(b) There are 10 types of diamond motifs, types 1,4,5,6,7,10 are coherent.



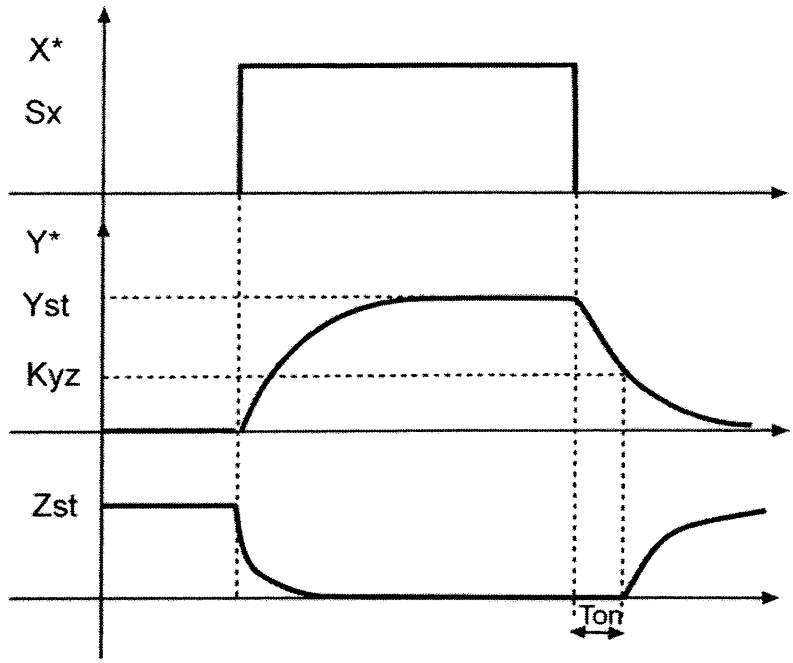
(c) A diamond generally has unequal response times for the arm through Y and the arm through Z. For example, suppose that Y and Z have the same production and degradation rates, but that their thresholds to activate W are different. Without loss of generality, suppose that Z has a lower threshold, $K_{zw}<K_{yw}$. We will solve for an AND gate at the W promoter. Following an ON step of S_x, both Y and Z must accumulate and cross their thresholds to activate W. The response is therefore governed by the higher of the two thresholds, since both Y and Z must cross their thresholds. Hence $T_{on} = 1/\alpha \log (Y_{st}/(Y_{st}-K_{yw}))$.

In contrast, after an OFF step, only one of the two regulators must go below its threshold to deactivate the AND gate at the W promoter. Again, the OFF time corresponds to the higher of the two thresholds because it is crossed first: $T_{off}=1/\alpha \log (Y_{st} / K_{yw})$.

The delay is asymmetric ($T_{ON} \neq T_{OFF}$) unless K_{yw} is equal to $Y_{st}/2$. Note that Z does not affect the dynamics at all in this circuit (assuming logic input functions).

4.5 *Type three.* Solve the dynamics of the type-3 coherent FFL (Fig 4.3) with AND-logic at the Z promoter in response to steps of S_x and S_y. Are there delays? What is the steady-state logic carried out by this circuit? Compare to the other coherent FFL types.

Solution:



Z is controlled by an AND gate over two repressors X and Y. Hence, both active X and active Y need to be below their corresponding thresholds to cause Z expression. An ON step of S_x causes X to become active as a repressor, resulting in an immediate stop in the production of Z and its exponential decay. Upon an OFF step of S_x X becomes inactive and Z production starts at a delay T_{on}, which is the time it takes for Y levels to decrease below the activation threshold K_{yz} :

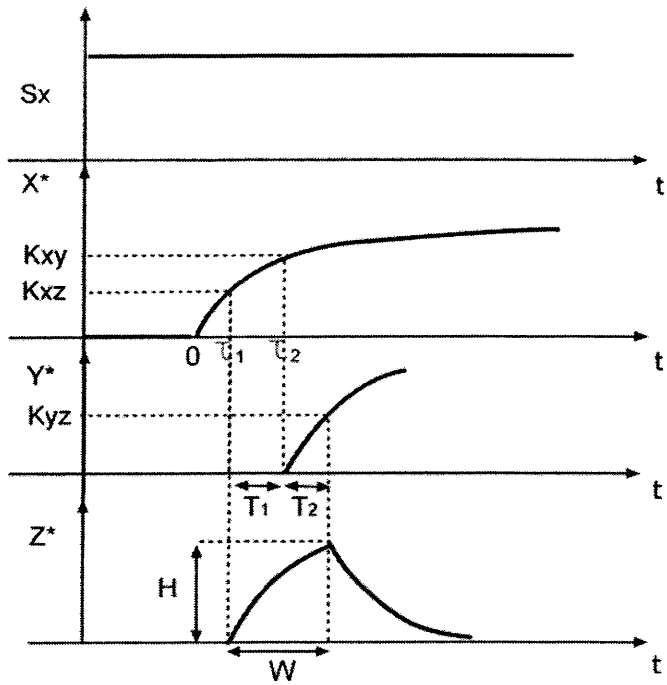
$T_{on}=1/\alpha \log(Y_{st}/K_{yz})$

Therefore this circuit has a sign-sensitive delay upon an OFF step of the input S_x.

4.6 *Shaping the pulse.* Consider a situation where X in an I1-FFL begins to be produced at time t=0, so that the level of protein X gradually increases. The input signal S_x is present throughout. a) How does the pulse shape generated by the I1-FFL depend on the thresholds K_{xz}, K_{xy} and K_{yz} and on β, the production rate of protein X? Analyze a set of genes $Z_1, Z_2, \ldots, Z_n$, all regulated by the same X and Y in I1-FFLs. b) Design thresholds such that the genes are turned ON in the rising phase of the pulse in a certain temporal order and turned OFF in the declining phase of the pulse with the same order. Design thresholds such that the turn-off order is opposite to the turn-on order. Plot the resulting dynamics.

Solution:

a)



As can be seen from the figure, the pulse width $W=T_1+T_2$ depends on K_{xy}, K_{xz} and K_{yz}. T_1, the delay from activation of Z to the activation of Y is $T_1=\tau_2-\tau_1$, where τ_1 is the time when $X^*=K_{xz}$ and τ_2 is the time when $X^*=K_{xy}$:

(4.6.1) $$(\beta_x/\alpha)(1-\exp(-\alpha\tau_1))=K_{xz} \Longrightarrow \tau_1=\frac{1}{\alpha}\log\left(\frac{1}{1-\alpha K_{xz}/\beta_x}\right)$$

similarly for τ_2:

(4.6.2) $$(\beta_x/\alpha)(1-\exp(-\alpha\tau_2))=K_{xy} \Longrightarrow \tau_2=\frac{1}{\alpha}\log\left(\frac{1}{1-\alpha K_{xy}/\beta_x}\right)$$

Finally:

(4.6.3) $$T_1=\tau_2-\tau_1=\frac{1}{\alpha}\log\left(\frac{1-\alpha K_{xz}/\beta_x}{1-\alpha K_{xy}/\beta_x}\right)$$

Let's consider two limits of the activation thresholds to see if this formula makes sense: if $K_{xz}= K_{xy}$, $T_1=0$, if $K_{xy}\longrightarrow\beta_x/\alpha$ then $T_1\longrightarrow\infty$.

T_2 is the time for Y to reach K_{yz}, the repression threshold of Z :

(4.6.4) $$T_2=\frac{1}{\alpha}\log\left(\frac{1}{1-\alpha K_{yz}/\beta_y}\right)$$

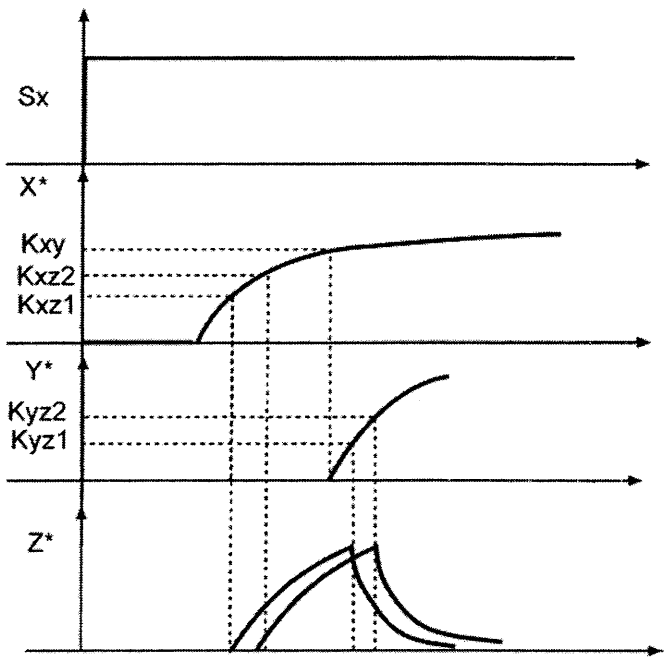
The pulse width is therefore:

(4.6.5) $$W=T_1+T_2=\frac{1}{\alpha}\log\left(\frac{1-\alpha K_{xz}/\beta_x}{(1-\alpha K_{xy}/\beta_x)(1-\alpha K_{yz}/\beta_y)}\right)$$

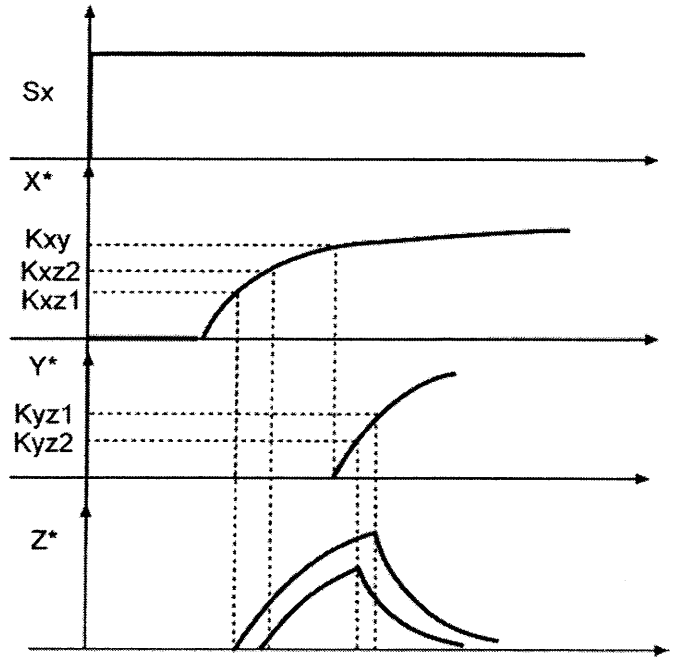
This can be intuitively understood when noting that increasing either K_{yz} and K_{xy} makes the pulse wider, whereas increasing K_{xz}, which determines the starting point of the pulse, makes the pulse shorter. The hight of the pulse H is the level of Z after a rise-time of W:

(4.6.6) $H=\beta_z/\alpha\ (1-\exp(-\alpha W))$

b) The following design on 2 output genes would result in a turn ON order equal to the turn OFF order: $K_{xz2} > K_{xz1}$ and $K_{yz2} > K_{yz1}$ (FIFO – First In First Out).



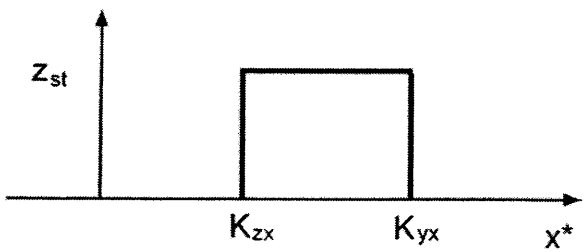
The design in which $K_{xz2} > K_{xz1}$ and $K_{yz2} < K_{yz1}$ would result in an opposite order for shutting down the pulse (LIFO – Last In First Out):



4.7 *Amplifying intermediate stimuli:* This problem highlights an additional possible function of incoherent type-1 FFLs for sub-saturating stimuli S_x. Consider an I1-FFL, such that the activation threshold of Z by X, K_{zx}, is smaller than the activation threshold of Y by X, K_{yx}. That is, Z is activated when $X^*>K_{zx}$, but it is repressed by Y when $X^*>K_{yx}$. Schematically plot the steady-state concentration of Z as a function of X*. Note that intermediate values of X* lead to the highest Z expression.

Solution

In the range $K_{zx}<X^*< K_{yx}$ The repression has no effect and the steady state level of Z will be high:

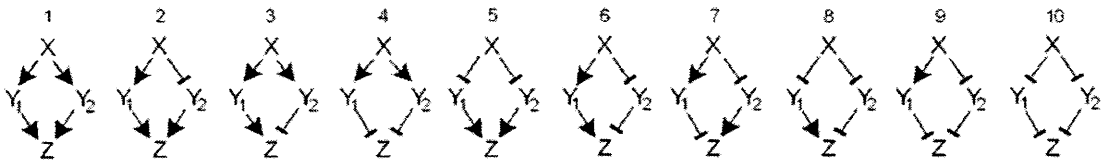


Our analysis of the I1 FFL assumed $X>K_{yx}$. In this regime the steady state is 0, but the temporal dynamics are shaped as a pulse.

4.8 *The diamond again.* The diamond pattern occurs when X regulates Y_1 and Y_2, and both Y_1 and Y_2 regulate gene Z. Analyze the 10 types of diamond

structures (where each edge is either activation + or repression -) with respect to their steady-state responses to the inputs S_x, S_{y1} and S_{y2}. Use an AND input function at the Z promoter. Do any diamond types lack responsiveness to any input? To all three inputs?

Solution



We can plot the truth table of all the configurations:

S_x,S_{y1},S_{y2}	1	2	3	4	5	6	7	8	9	10
0,0,0	0	0	0	1	0	0	0	0	1	1
1,0,0	0	0	0	1	0	0	0	0	1	1
0,1,0	0	0	0	1	0	0	0	1	1	0
1,1,0	0	0	1	0	0	1	0	0	0	1
0,0,1	0	0	0	1	0	0	1	0	0	0
1,0,1	0	0	0	0	0	0	0	0	1	1
0,1,1	0	0	0	1	1	0	1	0	0	0
1,1,1	1	0	0	0	0	1	0	0	0	1

Diamonds 1,3,4,5,8,9,10 are sensitive to all three inputs. Diamond 7 is not sensitive to S_{y1} (its steady state logic is: NOT(S_x) AND S_{y2}). Diamond 6 is not sensitive to S_{y2} (its steady state logic is: S_x AND S_{y1}). Diamond 2 always results in 0 (never on) therefore not sensitive to any input.

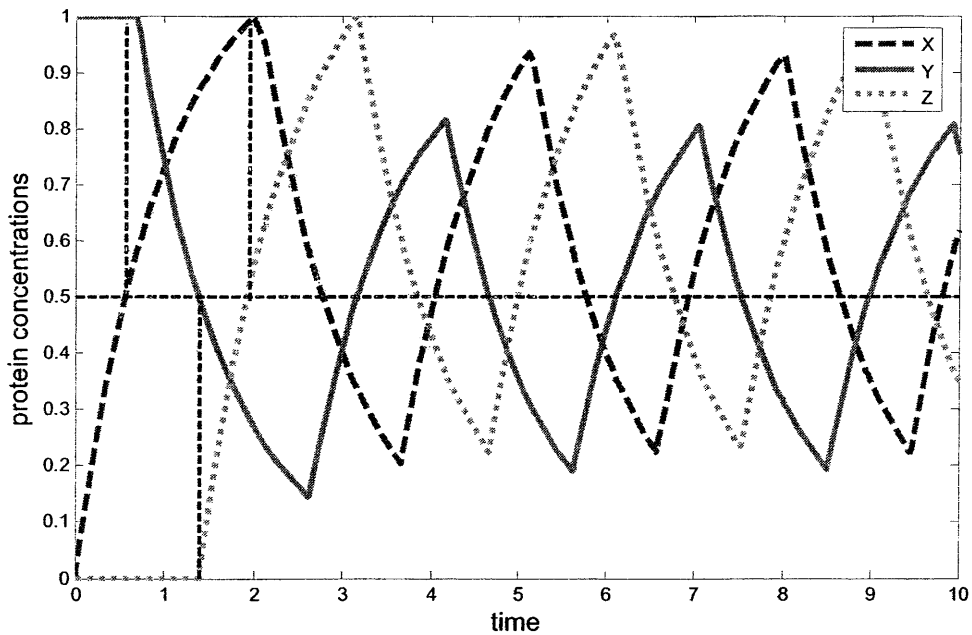
4.9 *Repressilator*. Three repressors are hooked up in a cycle X—|Y—|Z and Z—|X. What are the resulting dynamics? Solve graphically using logic input functions. This circuit was constructed in bacteria using three well studied repressors, one of which was also made to repress the gene for green-fluorescent protein (Elowitz and Leibler,

2000). What would the resulting bacteria look like under a microscope that dynamically records green fluorescence?

Solution:

Starting from an initial state where $X=0, Y=Y_{st}, Z=0$, X levels begin to rise until they reach K_{xy}. At this time, production of Y stops and Y exponentially decays. When $Y<K_{yz}$ Z production will start, and after Z reaches the level K_{zx} X production will stop. This cycle will continue oscillating in this fashion.

Results of Matlab simulation (time is in units of $1/\alpha$, $K_{xy}=K_{yz}=K_{zx}=0.5$, $\beta=1$):



Producing this plot involves defining the following MATLAB function:

```
function y=repressilator(t,x,options,alpha,beta,Kxy,Kyz,Kzx);

y(1)=beta*(x(3)<Kzx)-alpha*x(1);
y(2)=beta*(x(1)<Kxy)-alpha*x(2);
y(3)=beta*(x(2)<Kyz)-alpha*x(3);
y=y';
```

and then setting the parameters (alpha=1,beta=1,Kxy=Kyz=Kzx=0.5,Time=10)and running it with the line:

```
[t,y]=ode23('repressilator',[0 Time],[0 1 0],[],alpha,beta,Kxy,Kyz,Kzx)
```

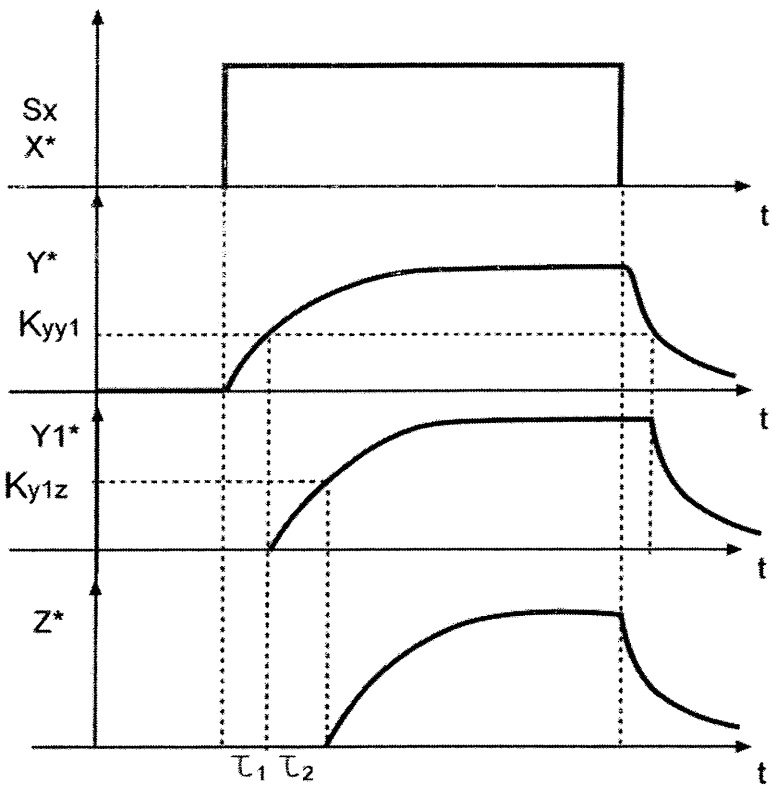
4.10 *Interconnected FFLs:* Consider a coherent type-1 FFL with nodes X, Y and Z, which is linked to another coherent type-1 FFL in which Y activates Y_1 which activates Z.

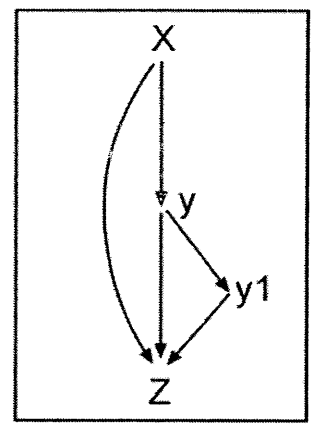
(a) Sketch the dynamics of Z expression in response to steps of the signals S_x, S_y and S_{y1}. Can the dynamics of the interconnected circuit be understood based on the qualitative behavior of each FFL in isolation?

(b) Repeat for the case where Y represses Z, so that the X, Y, Z FFL is an incoherent type-1 FFL. Assume that Y_1 binding to the Z promoter can alleviate the repressing effect of Y.

Solution:

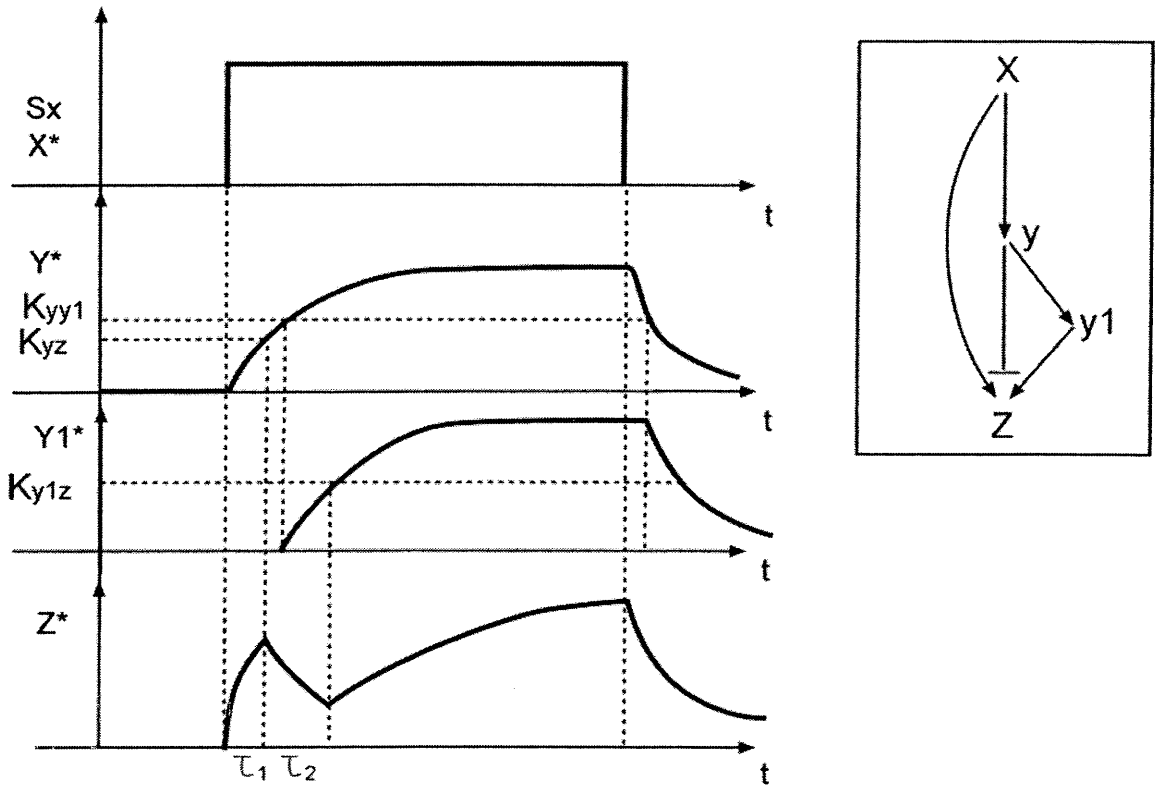
a)





Assuming an AND gate at the Z gene, this circuit is a sign-sensitive delay. The addition of the Y_1 gene adds a second delay τ_2, the time from Y_1 activation until it crosses its threshold for Z activation K_{y1z}. Upon shut down of the input stimulus S_x, Y and Z shut down immediately, whereas Y_1 shuts down in a delay. In summary, the behavior of the circuit output Z cannot be understood based on each FFL in isolation.

b)



In this case the Z dynamics upon a turn on step of S_x follows an 'oscillatory' rise to steady state, first increasing because Y levels are low, then after a delay τ_1 when y levels reach the threshold K_{yz} Z decreases, but after another delay τ_2 Y_1 crosses the thresholds upon which Y's repression on Z is alleviated, and Z increases again. The exact behavior depends on the circuit parameters (binding constants, α and β).

Exercises, chapter 5

5.1 *Robust timing.* Consider a SIM controlled by regulator X which activates downstream genes Z_i, i= 1 .. n genes with thresholds K_i. At time t=0, X begins to be produced at a constant rate β.

(a) Are there biological reasons that favor placing the thresholds K_i much smaller than the maximal level of X? Consider the case in which X is an activator that begins to be produced at time t=0, and consider the effects of cell-cell variations in the production rate of X.

(b) Would a design in which X is a repressor whose production stops at time t=0 provide more robust temporal programs? Explain.

Solution:

(a) Imagine that K_i is close to the maximal level of X, $X_{st} = \beta / \alpha$. The production rate of X, β, varies from cell to cell such that some cells have higher β than others over their entire generation time (Appendix D). Hence, there are cell-cell variations in X_{st}. In cells in which production is low, we might have $X_{st} < K_i$: in these cells X* does not cross K_i. Therefore the downstream gene Z_i is not expressed, and the cell is at a disadvantage. Thus, designs that provide the required timing and in which K_i is much smaller than the mean X_{st} (smaller than the lowest expected X_{st} given the noise in β), have an advantage.

Let's consider the case in which K_i are much lower than X_{st}. In this case, X crosses these thresholds at early times and can use the approximation of linear production $X(t) \sim \beta t$ (Eq. 2.4.7). Thus, the activation times of the genes, found by asking when $X(t)=K_i$, are

$t_i = K_i / \beta$.

Low thresholds thus ensure that all genes can be activated despite the noise in production rate. One can set K_i and β to achieve the required timing. Note that the activation times can vary from cell to cell if β varies. A factor 2 change in β would lead to a factor 2 change in t_i . The relative *order* of the turn-ON events of different genes in a SIM in the same cell would, however, not change.

(b) When X production stops, it decays $X(t)= (\beta / \alpha) \exp(-\alpha t)$. The turn ON time of gene *i* is the time when the level of repressor X goes below its threshold: $X(t_i) = K_i$, so that

$t_i = 1/ \alpha \log(\beta/ \alpha K_i)$.

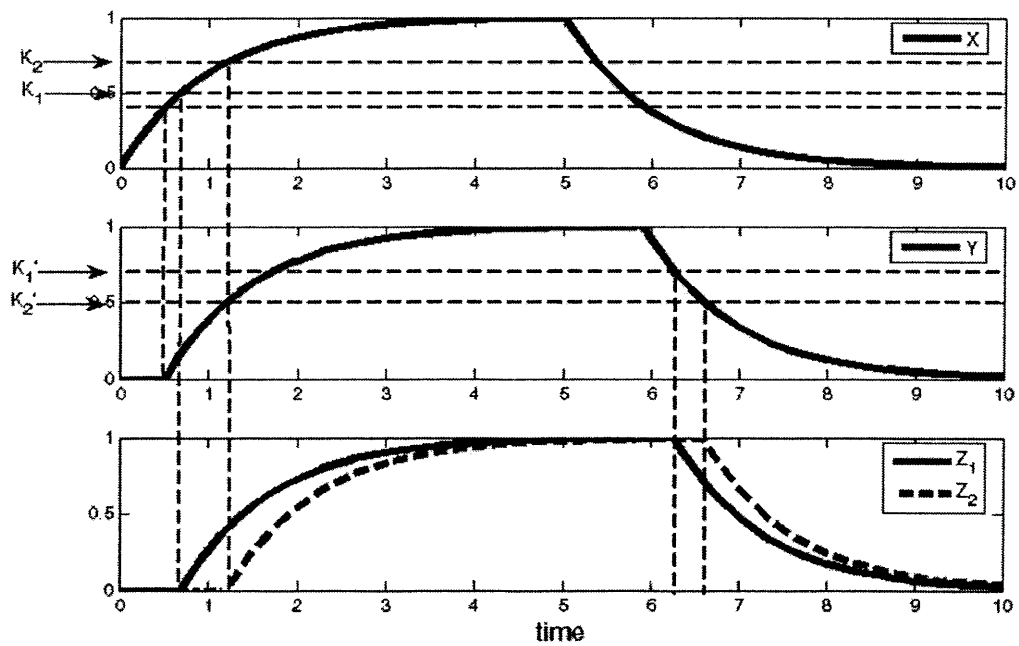
Note that β appears only in the logarithm in this expression, so that the activation time t_i is quite robust with respect to fluctuations in production rate, more so than for an activator whose concentration increases with time (where, as we saw above, $t_i \sim K_i / \beta$).

This robustness might be one reason that repressor cascades are often used in developmental transcription networks (chapter 6).

5.3 *The multi-output OR-FFL.* In a multi-output C1-FFL with OR-gate logic at the Z promoters, transcription-factor X begins to be produced at a constant rate β at time t=0. At time t=T, the production rate β suddenly becomes equal to zero. Calculate the dynamics of the downstream genes Zi. What are the delays between genes? (use logic input functions).

Solution:

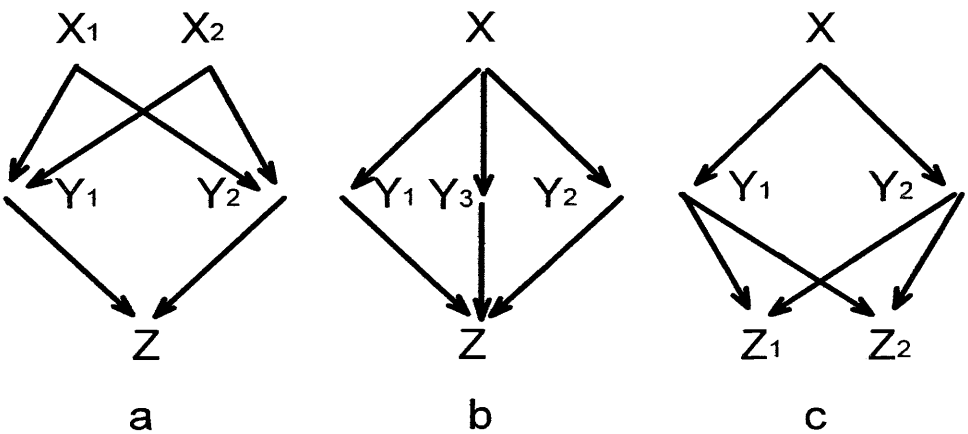
This circuit can be used to perform a FIFO (First In First Out) order of gene activation, by making $K_{XZ2}(K_2)>K_{XZ1}(K_1)$ and $K_{YZ2}(K_2')<K_{YZ1}(K_1')$.



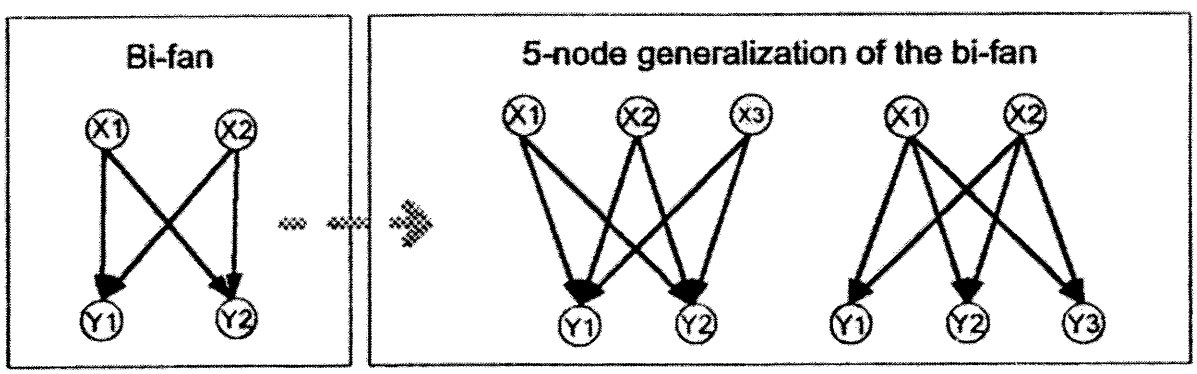
5.4 What are the topological generalizations of the diamond pattern ($X \rightarrow Y_1$, $X \rightarrow Y_2$, $Y_1 \rightarrow Z$, $Y_2 \rightarrow Z$) based on duplication of a single node and all of its edges? How are these different from DORs? What are the topological generalizations of the bi-fan ($X_1 \rightarrow Y_1$, $X_2 \rightarrow Y_1$, $X_1 \rightarrow Y_2$, $X_2 \rightarrow Y_2$)? Most of these 5-node generalizations of the diamond and bi-fan is a network motif in the neuronal network of *C. elegans* as we will see in the next chapter.

Solution:

The diamond pattern has three roles, and can be generalized by replicating the X role (a) Y role (b) or Z role (c). All of these motifs have a delay of 2 edges between the input X and the output Z as opposed to a direct connection from input to output in a DOR.



The 5-node generalizations of the bi-fan network motif are:



(from Kashtan et al, Phys Rev E 2004)

Exercises, chapter 6

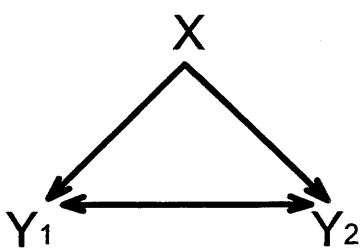
6.1 *Memory in the regulated-feedback network motif:* Transcription factor X activates transcription factors Y_1 and Y_2. Y_1 and Y_2 mutually activate each other. The input function at the Y_1 and Y_2 promoters is an OR-gate. At time t=0, X begins to be produced from an initial concentration of X=0. All production rates are β=1 and degradation rates are α=1. All of the activation thresholds are K=0.5. At time t=3, X production stops.

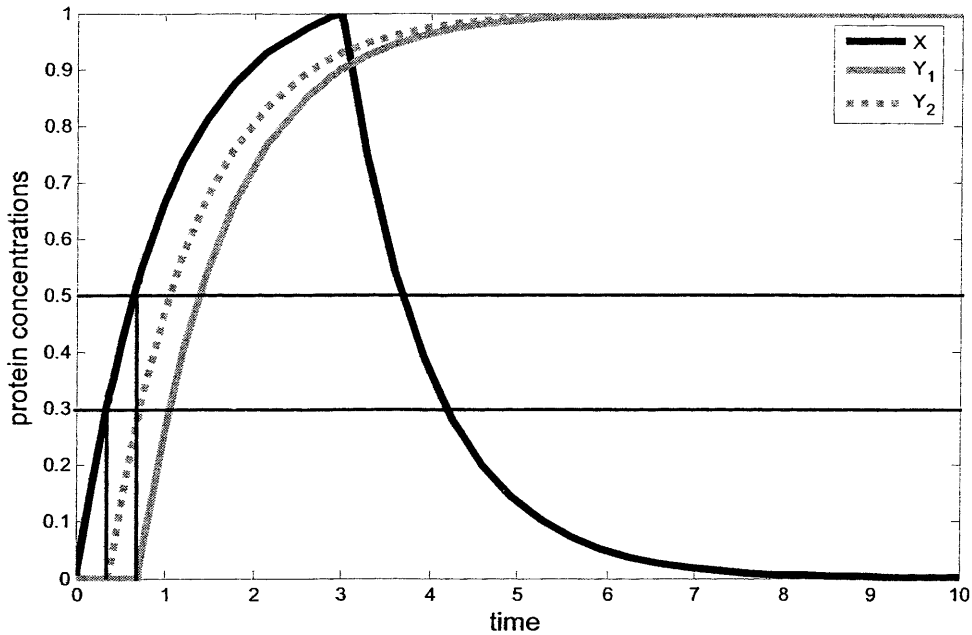
(a) Plot the dynamics of X, Y_1 and Y_2. What happens to Y_1 and Y_2 after X decays away?

(b) Consider the same problem, but now Y_1 and Y_2 repress each other, and X activates Y_1 and represses Y_2. Y_1 is activated if either X binds it or Y_2 does not bind it. Y_2 is activated only if neither X or Y_1 bind it. At time t=0, X begins to be produced, and the initial levels are X=0, Y_1=0 and Y_2=1. At time t=3, X production stops. Plot the dynamics of the system. What happens after X decays away?

Solution:

a)





Note that the levels of Y_1 and Y_2 are locked in an on state even long after X production has stopped.

The matlab function for simulating this circuit is :

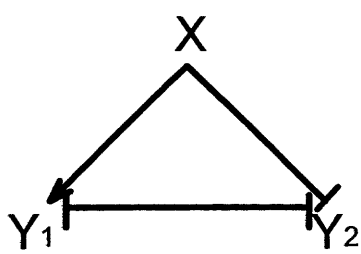
```
function y=mutually_regulating(t,x,options,alpha,beta,Kxy,Kxz,Kyz,Kzy);

y(1)=beta*(t<3)-alpha*x(1);
y(2)=beta*((x(1)>Kxy)|(x(3)>Kzy))-alpha*x(2);
y(3)=beta*((x(1)>Kxz)|(x(2)>Kyz))-alpha*x(3);
y=y';
```

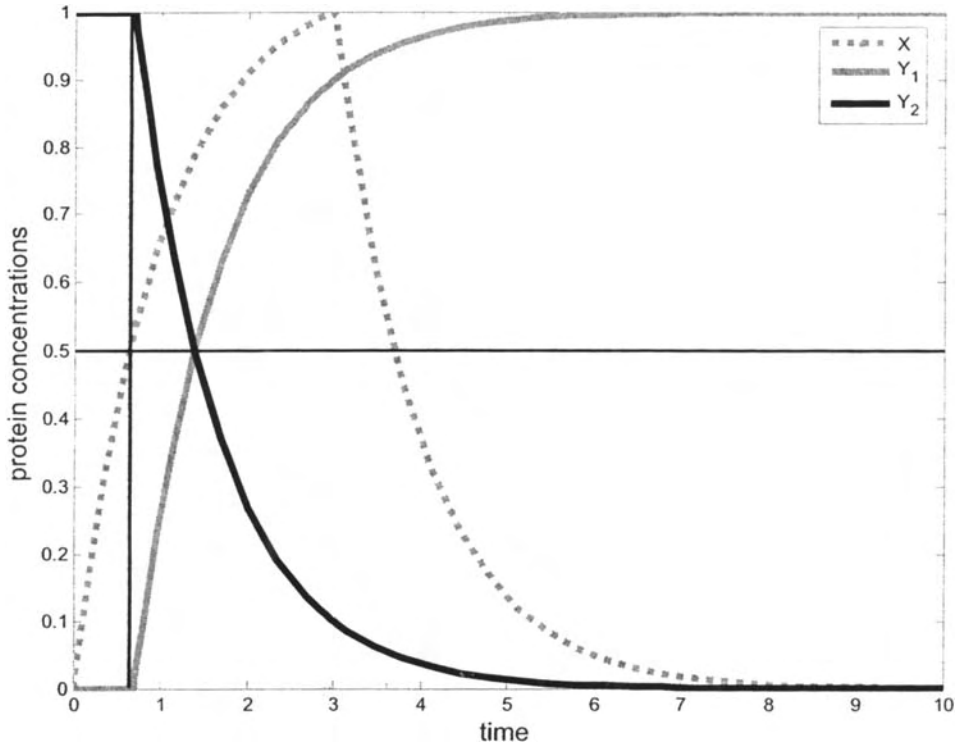
and should be run with the command :

```
[t,y]=ode23('mutually_regulating',[0 Time],[0 0 0],[],alpha,beta,Kxy,Kxz,Kyz,Kzy);
```

In the simulation run shown in the figure K_{xy1}=0.5, K_{xy2}=0.3.

b)





Here we assume the logic function on Y_1 is : $Y_1=(X_1>K_{XY1})$ OR $(Y_2<K_{Y1Y2})$,
And the logic on Y_2 is $Y_2=(X_1<K_{XY2})$ AND $(Y_1<K_{Y2Y1})$. The circuit is locked in a state where Y_1 levels are high and Y_2 levels are low.
Both of the circuits in a) and b) act as memory elements. Their state remains constant long after the circuit input S_x has passed. (In the simulation run shown in the figure $K_{XY1}=K_{XY2}=0.5$).

6.2 *Kinases with double phosphorylation:* Kinase Y is phosphorylated by two input kinases X_1 and X_2 which work with first-order kinetics with rates v_1 and v_2. Y needs to be phosphorylated on two sites to be active. The rate of phosphorylation and rate of de-phosphorylation of the two phosphorylation sites on Y are the same. Derive the input function, the fraction of doubly phosphorylated Y as a function of the activity of X_1 and X_2.

Solution:

The kinase Y exists in three states, with zero, one and two phosphorylations, denoted Y_o, Y_1 and Y_2. The total amount of Y is conserved

(6.2.1) $Y_o+Y_1+Y_2=Y.$

The rate of change of Y_1 is given by an equation that balances the rate of the input kinases and the action of the phosphatases, taking into account the flux from Y_0 to Y_1 and from Y_1 to Y_2, as well as de-phosphorylation of Y_2 to Y_1:

(6.2.2) $dY_1/dt = V_1 X_1 Y_0 + V_2 X_2 Y_0 - V_1 X_1 Y_1 - V_2 X_2 Y_1 - \alpha Y_1 + \alpha Y_2$

And the dynamics of Y_2 is

(6.2.3) $dY_2 / dt = v_1 X_1 Y_1 + v_2 X_2 Y_1 - \alpha Y_2$

At steady state, $dY_2 / dt = 0$, and Eq 6.2.3 yields

(6.2.4) $(v_1 X_1 + v_2 X_2) Y_1 = \alpha Y_2$

using the weights $w_1 = v_1/\alpha$ and $w_2 = v_2/\alpha$,

(6.2.5) $Y_1 = Y_2 / (w_1 X_1 + w_2 X_2)$

Summing Eq 6.2.3 and 6.2.2 yields $d(Y_1 + Y_2)/dt = Y_0 (v_1 X_1 + v_2 X_2) - \alpha Y_1$, so that at steady-state

(6.2.6) $Y_0 = Y_1/(w_1 X_1 + w_2 X_2) = Y_2 / (w_1 X_1 + w_2 X_2)^2$

Using Eq 6.2.1, we find

(6.2.7) $Y = Y_0 + Y_1 + Y_2 = (1 + 1/u + 1/u^2) Y_2$

where

(6.2.8) $u = w_1 X_1 + w_2 X_2$

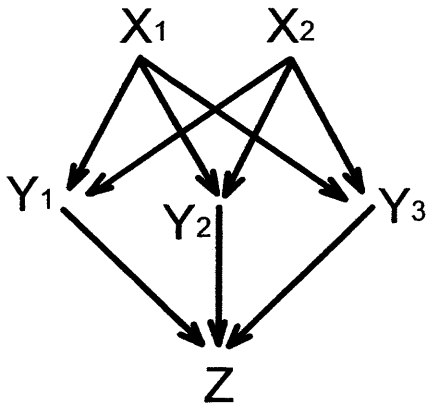
Thus, the desired input function is

(6.2.9) $Y_2/Y = u^2/(u^2 + u + 1)$

Note that for n phosphorylations, the input function is $Y_n / Y = u^n / (1 + u + \ldots + u^n)$

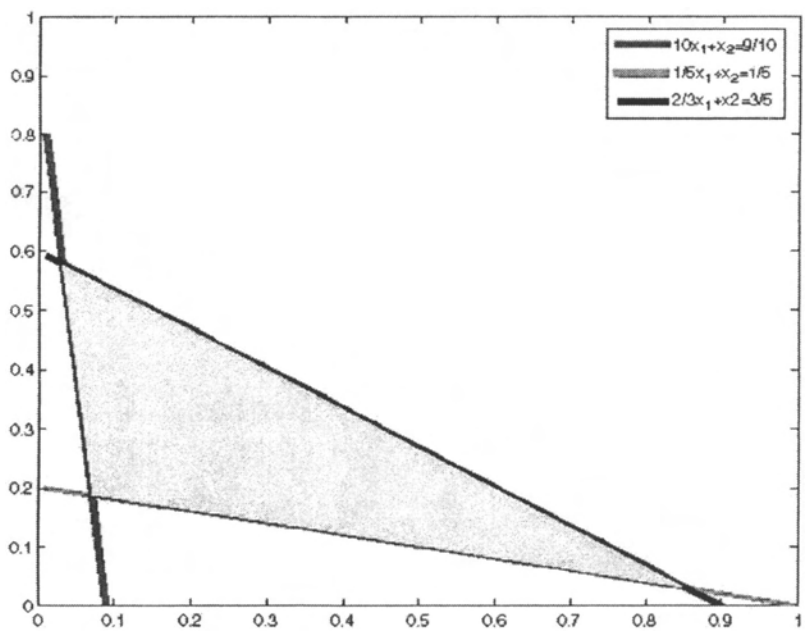
6.3 Design a multi-layered perceptron with two input nodes, one output node and as many intermediate nodes as needed, whose output has a region of activation in the shape of a triangle in the middle of the X_1-X_2 plane.

Solution:



The first layer parameters are displayed in the figure legend as $W_{1j}X_1+ W_{2j}X_2=T_j$, where W_{ij} are the weights connecting X_i to Y_j and T_j is the threshold of perceptron j. For Y_1 : $W_{11}=10$, $W_{21}=1$, $T_1=9/10$, For Y_2 : $W_{12}=1/5$, $W_{22}=1$, $T_1=1/5$, For Y_3 : $W_{13}=2/3$, $W_{21}=1$, $T_1=3/5$.

The weights of the second layer are : $W_{Y1Z}=2/3$,$W_{Y2Z}=2/3$, $W_{Y3Z}= -2$, and $T_Z=1$.



6.4 *Dynamics of a protein-kinase cascade:* Protein kinases X_1, X_2, ...X_n act in a signaling cascade, such that X_1 phosphorylates X_2 which, when phosphorylated, acts to phosphorylate X_3, etc.

(a) Assume sharp activation function. What is the response time of the cascade, the time from activation of X_1 to a 50% rise in the activity of X_n?

(b) What is the effect of the kinase rates on the response time? Of the phosphatase rates? Which have a larger effect on the response time? (Heinrich et al., 2002)

Solution:

(a) The rate of change of active (phosphorylated) X_i are given by the difference between the sharp phosphorylation rate by kinase X_{i-1}, with rate v_{i-1}, and the de-phosphorylation process by the phosphatases that work on X_i at rate α_i.

(6.4.1) $$dX_i/dt = v_i\, \theta(X_{i-1}>K_{i-1}) - \alpha_i X_i$$

thus, X_i begins to increase at the time that X_{i-1} crosses its threshold K_{i-1}. At this point, X_i begins to increase with the familiar exponential convergence to steady-state (e.g. Eq 2.4.6):

(6.4.2) $$X_i = (v_i/\alpha_i)\ [\,1 - \exp(-\alpha_i (t-t_i))\,]$$

When the concentration of the kinase X_i (in its phosphorylated form) crosses the activation threshold, it begins to activate the next kinase in the cascade. Thus the onset of phosphorylation of X_{i+1}, denoted t_{i+1}, can be found by solving

(6.4.3) $$K_i = (v_i/\alpha_i)\,[1-\exp(-\alpha_i (t-t_i))\,]$$

yielding

(6.4.4) $$t_{i+1}=t_i + 1/\alpha_i \log(\,1/(1 - \alpha_i K_i / v_i)$$

We thus find that

(6.4.5) $$t_n = \Sigma_i\, 1/\alpha_i \log(1/(1 - \alpha_i K_i / v_i))$$

(b) According to Eq 6.4.5, the phosphatase rates α_i have a large effect on the response times. If these rates are very different for each kinase in the cascade, the response time is dominated by the slowest rate, because it has the largest $1/\alpha_i$. In contrast to the strong dependence of phosphatase rates, the response time is only weakly affected by the kinase velocities v_i, because they appear inside the logarithm in Eq 6.4.5.

6.5 Dynamics of a linear protein kinase cascade(Heinrich et al., 2002): In the previous problem, we analyzed the dynamics of a cascade with sharp input functions. Now, we consider the case where the kinetics of the kinases are zero-order, that is when the activated upstream kinase is found in much smaller concentrations that its

un-phosphorylated target. In zero-order kinetics, the rate of phosphorylation depends only on the upstream kinase concentration and not on the concentration of its substrate. In this case, we need to analyze a *linear* set of equations

(6.5.1) $dX_i/dt = v_{i-1} X_{i-1} - \alpha_i X_i$

The **signal amplitude** is defined by

(6.5.2) $$A_i = \int_0^\infty Xi(t)\, dt$$

and the **signal duration** by

(6.5.3) $$\tau_i = \int_0^\infty t\, Xi(t)\, dt / A_i$$

In many signaling systems, the duration of the signaling process is important, in the sense that brief signals can sometimes activate different responses than prolonged signals.

(a) The cascade is stimulated by a pulse of X_1 activity with amplitude A_1, that is $\int_0^\infty X_1(t)\, dt = A_1$. What is the amplitude of the final stage in the cascade, A_n?

(b) What is the signal duration of Xn?

(c) How do the kinase and phosphatases rates affect the amplitude and duration of the signal? Compare to exercise 6.4.

Solution:

To find the amplitude, let's take an integral over time of both sides of Eq 6.5.1.

(6.5.4) $$\int_0^\infty dt\, dX_i/dt = \int_0^\infty v_{i-1} X_{i-1}\, dt - \int_0^\infty \alpha_i X_i\, dt$$

Note that the integral in the left hand side is equal to $X_i(\infty)-X_i(0)$. Now, because the signal begins at t=0 and decays at long times, we have $Xi(0)=Xi(\infty)=0$. The integrals on the right hand side give rise to amplitudes as defined in Eq P6.5.2:

(6.5.5) $0 = v_{i-1} A_{i-1} - \alpha_i A_i$

Thus,

(6.5.6) $A_i = (v_{i-1}/ \alpha_i) A_{i-1}$

Therefore, by induction, we find that the amplitude is the product of kinase velocities divided by the product pf phosphatase rates

(6.5.7) $A_n = (v_{n-1}/ \alpha_n) A_{n-1} = (v_{n-1} v_{n-2}/\alpha_n \alpha_{n-1}) A_{n-2} = \ldots$

$$= (v_{n-1}\, v_{n-2} \ldots v_2\, v_1 / \alpha_n\, \alpha_{n-1} \ldots \alpha_2)\, A_1$$

(b) to find the signal duration, we take an integral over time of the dynamic equation 6.5.1 to find

$$(6.5.8) \qquad \int_0^\infty dt\, t\, dX_i/dt = \int_0^\infty dt\, v_{i-1}\, t\, X_{i-1} - \int_0^\infty dt\, \alpha_i\, t\, X_i$$

The left-hand-side integral can be solved using integration by parts, to yield

$$(6.5.9) \qquad \int dt\, t\, dX_i/dt = [t\, X_i]_0^\infty - \int X_i\, dt = -A_i$$

The right hand side is just proportional to the durations of X_{i-1} and X_i

So that we find

$$(6.5.10) \qquad -A_i = v_{i-1}\, \tau_{i-1}\, A_{i-1} - \alpha_i\, \tau_i\, A_i$$

Hence, we have, dividing both sides by A_i and using Eq 6.5.6 to eliminate A_{i-1},

$$(6.5.11) \qquad 1 = \alpha_i\, \tau_i - \alpha_i\, \tau_{i-1}$$

Which can be rearranged to yield

$$\tau_i - \tau_{i-1} = 1 / \alpha_i$$

Hence, the signal duration of the final step in the cascade is just the sum over the reciprocal phosphatases rates

$$\tau_n = \Sigma_i\, 1 / \alpha_i$$

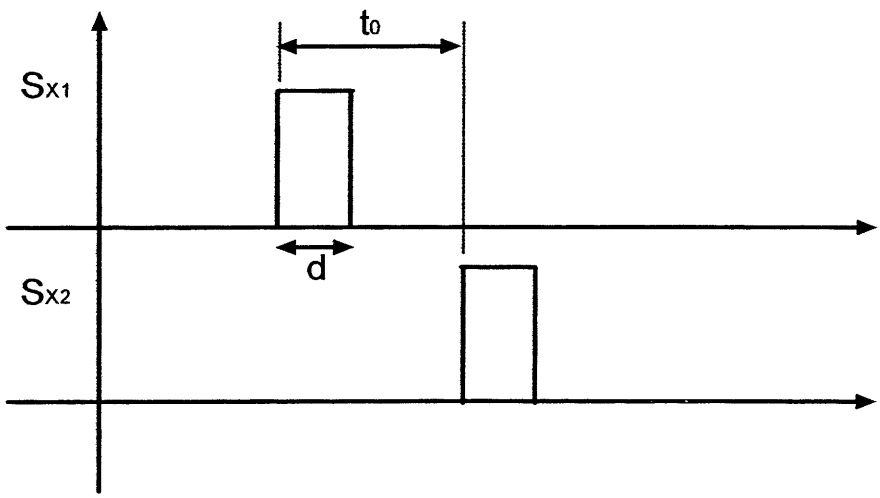
(c) We have just found that phosphatases rates α_i affect both amplitude and duration in zero-order kinetics cascades. The larger the phosphatase rates, the smaller the amplitude and the shorter the duration. In contrast, the kinase rates do not affect duration at all, and affect the signal amplitude proportionally. This is similar to problem 6.4, where we saw that phosphatase rates affect timing much more strongly than kinase velocities. In both models, the sum over $1 / \alpha_i$ determined the timing. This principle is identical to that which we saw in transcription networks, whose response times are governed inversely by the degradation/dilution rates: these rates are the eigenvalues of the dynamic equations. The strong effect of phosphatases on signal duration and the weak effect of kinases was demonstrated experimentally (see experiments cited in (Hornberg et al., 2005)).

6.6 *Coincidence detection:* Consider the two-input FFL motif of Fig 6.22. The two inputs receive brief activation pulses at a slight delay. The pulse of S_{x1} has duration d. At time t_0 after the start of the pulse, a pulse of S_{x2} begins and lasts for duration d.

(a) What is the minimal S_{x1} input pulse duration d that can activate Z without need for the second pulse of S_{x2}?

(b) Plot the region in which Z shows a response on a plane whose axes are pulse duration d and inter-pulse spacing t_0.

Solution:



(a) Let's denote Y's threshold for Z activation by K. In order for S_{x1} input to suffice we demand that:

(6.6.1) $\beta/\alpha\ (1-\exp(-\alpha d))>=K$

The minimal pulse width is thus :

(6.6.2) $$d = 1/\alpha \log(\frac{1}{1-K\alpha/\beta})$$

(b) If $t_0<d$ we effectively have one pulse of length $d'=d+t_0$, and using (6.6.2) we get:

$$d > 1/\alpha \log(\frac{1}{1-K\alpha/\beta}) - t_0$$

We shall now solve for the case where $t_0>d$. At the end of the first pulse Y levels reach:

(6.6.3) $$Y_0 = \beta/\alpha(1-e^{-\alpha d})$$

At this stage Y levels exponentially decay. At the beginning of the second pulse Y level is:

(6.6.4) $$Y_0 = \beta/\alpha(1-e^{-\alpha d})e^{-\alpha(t_0-d)}$$

After the second pulse Y's dynamics follows the familiar form:

(6.6.5) $Y=Y_0 + (\beta/\alpha - Y_0)(1-\exp(-\alpha t))$

By the end of the second pulse, Y's levels reach:

(6.6.6) $Y=Y_0 + (\beta/\alpha-Y_0)(1-\exp(-\alpha d))=Y_0 \exp(-\alpha d)+\beta/\alpha(1-\exp(-\alpha d))$

Inserting (6.6.4):

(6.6.7) $Y=\beta/\alpha(1-e^{-\alpha d})e^{-\alpha t0}+\beta/\alpha(1-e^{-\alpha d})= \beta/\alpha(1-e^{-\alpha d})(1+e^{-\alpha t0})$

Z will respond as long as the level of Y is greater then K at the end of the second pulse :

(6.6.8) $\beta/\alpha\ (1-\exp(-\alpha d))\ (1+\exp(-\alpha t_0))>=K$

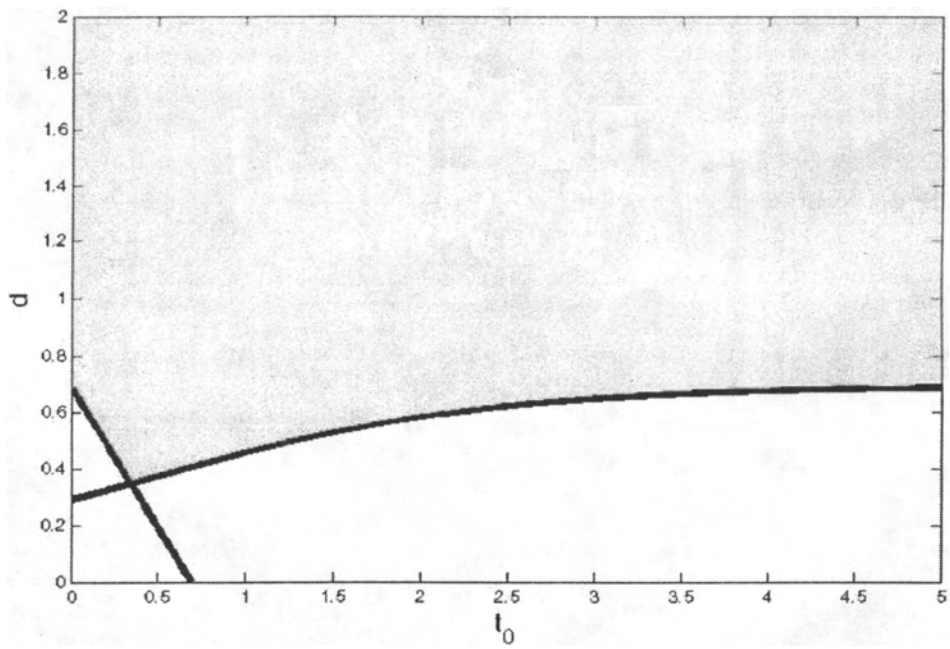
or equivalently:

(6.6.9) $\beta/\alpha\ (1-\exp(-\alpha d))\ >=K'(t_0)=K/(1+\exp(-\alpha t_0))$

solving for d as a function of t_0:

$$d(t_0) = 1/\alpha \log(\frac{1}{1-K'(t_0)\alpha/\beta}) = 1/\alpha \log(\frac{1}{1-K\alpha/(\beta*(1+e^{-\alpha t_0}))})$$

Let's graphically plot this function for $\alpha=\beta=1$ and K=1/2:



Note that as $t_0 \longrightarrow \infty$ the first pulse has no effect and we return to the solution of (6.6.2) for a single pulse (the d asymptote). As t_0 becomes shorter the pulse width can be shorter, as the basal level Y_0 at the beginning of the second pulse are higher.

6.7. Consider the diamond generalization (question 5.4) that has two input X_1 and X_2 and a single output Z. This two-layer perceptron pattern has 6 edges. Assume that all neurons are 'integrate and fire', and each can have a threshold of K=1 or K=-1. Assume that neurons have voltage 0, unless the weighted inputs exceed K, in which

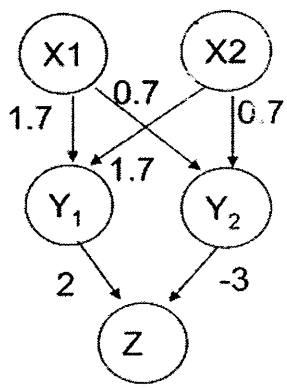
case they assume voltage 1.Weights on the edges can be positive or negative real numbers.

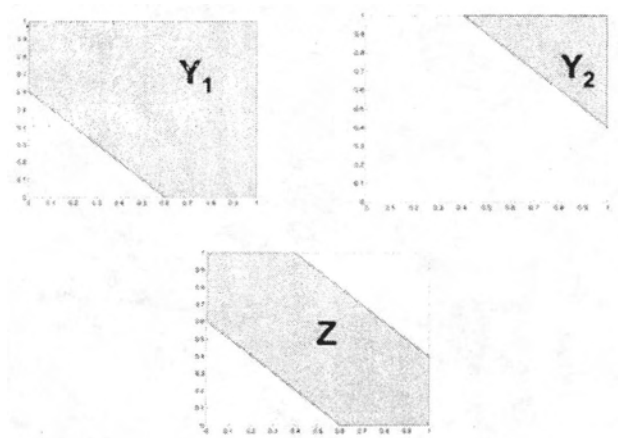
(a) Design weights such that this circuit computes the XOR (exclusive-or) function, where Z=1 if either X_1=1 or X_2=1 but Z=0 if both X_1=1 and X_2=1. This function is denoted Z=X_1 XOR X_2.

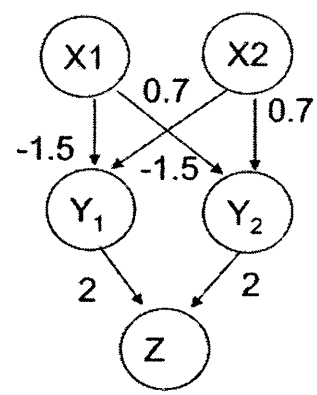
(b) Design weights such that his circuit computes the 'equals' function, in which Z=1 only if X_1 and X_2 are the same (both 0 or both 1), and Z=0 otherwise (that is, Z=X_1 EQ X_2).

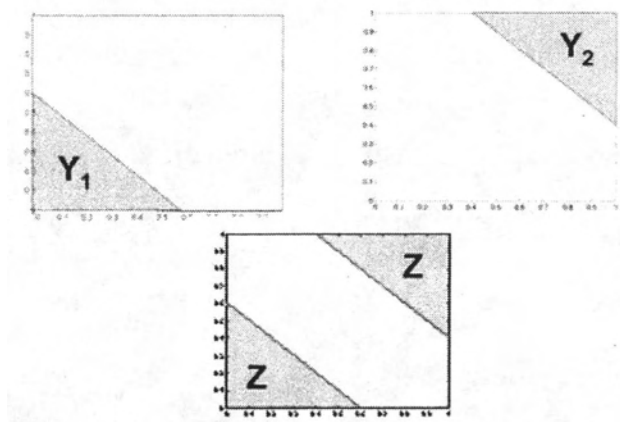
Solution:

The solution to the XOR problem is implemented with the weights (1.7,1.7) for Y_1 and (0.7,0.7) for Y_2. All thresholds are 1. The solution to the EQ problem is implemented with the weights (-1.5,-1.5) for Y_1 and (0.7,0.7) for Y_2. The Y_1 threhold is -1, and Y_2 threshold is 1. The second layer performs an OR gate on Y_1 and Y_2 at the Z perceptron with weights (2,2) and threshold 1.









Exercises, chapter 7

7.1 *Robust model with two methylation sites*: The receptor X can be methylated on two positions, and can thus have 0, 1 or 2 methyl groups, denoted Xo, X_1 and X_2. The enzyme R works at saturation (zero-order kinetics) to methylate Xo and X_1. The de-methylating enzyme B works only on the active receptor conformation, removing methyl groups with equal rate from X_1* and X_2*. For simplicity, assume that B works with first-order kinetics. The reactions are:

Xo→X_1 at rate R V_R Xo/(X_1+Xo), the last factor because R is distributed between Xo and X_1

X_1→X_2 at rate R V_R X_1/(X_1+Xo)

X_1←→X_1* rapid transitions at rate that depends on ligand level

X_2←→X_2* rapid transitions at rate that depends on ligand level.

X_1*→Xo rate B V_B X_1*

X_2*→X_1 rate B V_B X_2*

(a) What is the steady state activity A=X_1*+X_2*? Does it depend on ligand l? Is there exact adaptation?

(b) Estimate the adaptation time, the time needed for 50% adaptation after addition of saturating attractant. Note that to adapt to saturating attractant, virtually all of the receptors need to be doubly methylated.

(c) Spudich and Koshland found that different cells in a population have different steady-state activity and different adaptation-times (Non-genetic individuality: chance in the single cell. Nature 262:467 (1976)). Moreover, these two features were found to be correlated: The higher the activity A, the shorter the adaptation time τ in a given cell, with A~1/τ. Explain this finding using the model, based on cell-cell variations in the concentration of R . (Barkai and Leibler, 1997).

Solution:

(a) The rate of change of the doubly methylated receptor concentration and the non-methylated receptor concentration are:

(7.1.1) $dX_2/dt = R\ V_R\ X_1/(X_1+Xo) - B\ V_B\ X_2^*$

(7.1.2) $dXo/dt = -R\ V_R\ Xo/(Xo+X_1) + B\ V_B\ X_1^*$

Subtracting these two equations, yields

(7.1.3) $dX_2/dt - dX_0/dt = R\,V_R - B\,V_B\,(X_1^* + X_2^*) = R\,V_R - B\,V_B\,A$

at steady-state, the activity $A = X_1^* + X_2^*$ is

(7.1.4) $A_{st} = R\,V_R / B\,V_B$

This activity does not depend on the ligand level l. Therefore this mechanism generates exact adaptation.

(b) In the case of saturating ligand, all receptors in all of their forms bind attractant ligand. The attractant reduces the activity of all methylated receptors, and thus at initial times X_1^* is small. In addition, when adaptation is completed, X_1^* is small, because the majority of receptors need to be doubly methylated in order to balance the strong inhibitory effect of the saturating attractant. Hence, X_1^* is relatively small throughout most of the dynamics. Since X_1^* is small, the de-methylation flux from X_1^* to Xo is small. Hence, to a good approximation, Xo dynamics reflect only a reduction due to the action of R and the term with B in 7.1.2 is negligible:

(7.1.5) $dXo/dt \approx -R\,V_R\,Xo / (Xo + X_1)$

so that Xo drops with time. At initial times (before attractant addition) let's denote by q the fraction of Xo among the possible substrates of CheR, $q = Xo/(Xo+X_1)$. Thus, the initial slope of the drop in Xo is $-\,q\,R\,V_R$. The adaptation time to saturating ligand (time to recover to 50% activity) is the time needed to build enough methylated receptors to restore activity, at the expense of most of the un-methylated ones. Thus, it is approximately the time for Xo to decline to 50% of Xo. This adaptation time is equal to the number of methylation reactions needed (that is, methylations equal to 50% of Xo) divided by the rate at which they occur, namely (ignoring the changes in q over this time):

(7.1.6) $\tau \sim 0.5\,Xo / q\,R\,V_R$

Thus, the adaptation time becomes shorter the more R enzymes exist in the cell. This makes sense, because the more R enzymes, the faster methylation occurs and the faster the adaptation.

Note that the single-methylation model discussed in the text has a different adaptation time, governed by B and not R. This is because we can not ignore the flux from X_1^* to Xo, that is necessary to produce exact adaptation in the single-methylation model. But B governs the adaptation time only if we restrict ourselves to a single methylation site, as we did for simplicity and clarity in the text. In reality there are 4-5 sites. The adaptation time is generally governed by R in models with

more than one methylation site (Barkai and Leibler, 1997). In experiments, the adaptation time is found to decrease with R (Alon et al., 1999), in agreement with the multi-site models.

(c) We saw above that the adaptation time varies as $\tau \sim 1/R$ (Eq. 7.1.6) , and that the steady-state activity varies as $A_{st} \sim R$ (Eq. 7.1.4). Thus, if R is the protein with largest variation between genetically identical cells, one would expect that $A_{st} \sim 1/\tau$, as observed. The protein R is the least abundant chemotaxis signaling protein in *E. coli*, with on the order of 100 copies per cell, whereas there are on the order of several thousand copies of CheB, CheY, CheZ and CheA per cell. CheR may therefore be the most prone to stochastic variations.

7.2 *Integral feedback:* A heater heats a room. The room temperature is T. The temperature increase in proportion to the power of the heater, P, to other sources of heat S, and to thermal diffusion to the outside at a rate proportional to T

(7.2.1) $dT / dt = a P + S - b T$

An integral feedback device is placed in order to keep the room temperature at a desired point T_o. In this feedback loop, the power to the heater is proportional to the integral over time of the error in temperature, $T-T_o$

(7.2.2) $$P = P_0 - k \int (T - T_0) dt$$

This feedback loop thus reduces the power to the heater if the room temperature is too high $T>T_o$, and increases the power when the room temperature is too low. Taking the time derivative of the power, we find

(7.2.3) $d P / d t = -k (T-T_o)$

(a) Show that the steady-state temperature is T_o, and that this steady-state does not depend on any of the system parameters including the rooms' thermal coupling to the heater *a*, the additional heat sources S and the rooms' thermal coupling with the outside *b*, or the strength of the feedback *k*. In other words, integral feedback shows robust exact adaptation of the room temperature.

(b) Demonstrate that integral feedback is the *only* solution that shows robust exact adaptation of the room temperature, out of all possible linear control systems. That is, assume a general linear form for the controller

(7.2.4) $dP / dt = c_1 T + c_2 P + c_3$

and show that integral feedback as a structural feature of the system is necessary and sufficient for robust exact adaptation.

Solution:

(a) At steady state dT/dt=0 and dP/dt=0. From 7.2.3 we get:

$dP/dT = -k(T-T_0) = 0$

hence

$T=T_0$ irrespective of any other system parameters.

(b) For the general form:

$dP / dt= c_1 T +c_2 P + c_3$

we have two possibilities :

(1) $c_2 \neq 0$. In this case at steady state from 7.2.4:

$P_{st}=-(c_1/c_2)T - c_3/c_2$

and from 7.2.1:

$T_{st}=(a/b)P_{st}+S/b = -(ac_1/bc_2)T_{st}-(ac_3/bc_2)+S/b$

Solving for the steady state temperature:

$T_{st}[1+(ac_1/bc_2)]=S/b-(ac_3/bc_2) \Longrightarrow T_{st}= [S/b-(ac_3/bc_2)]/ [1+(ac_1/bc_2)]$

Which depends on many of the system parameters (such as room heat capacity a and external sources S).

(2) $c_2=0$. In this case $T_{st}=-c_3/c_1$ and does not depend on any system parameters. We can now write equation 7.2.4 as :

$dP / dt= c_1 T + c_3 = c_1(T-T_{st})$

which is the integral feedback form (7.2.3).

7.3 *Integral feedback in chemotaxis:* Demonstrate that a simple linear form of the robust toy model for chemotaxis contains integral feedback. What is the integrator in this biological system? (Yi et al., 2000)

Solution: In a linear model, CheR works at saturation and CheB works with first-order kinetics, and only on the active receptors. The rate of change of the total number of methylated receptors ($X_{m,t}=X_m+X_m^*=X_m+A$) is given by the difference between the methylation and de-methylation rates:

(7.3.1) $\quad d X_{m,t} / d t =V_R R -V_B B A$

This can be rewritten in terms of the difference between the activity A and its steady-state value A_{st}

(7.3.2) $d X_{m,t} / d t = -V_B B (A-A_{st})$

where the steady-state activity is

(7.3.3) $A_{st} = V_R R / V_B B.$

The total number of methylated receptors $X_{m,t}$ thus acts as the integrator in the system that integrates the error in activity over time

(7.3.4) $X_{m,t} \sim -V_B B \int (A - A_{st}) dt$

The activity A is analogous to the room temperature in problem 7.2. To complete the analogy with problem 7.2, let's write a detailed equation for the rate of change of activity $A=X_m^*$. The number of methylated active receptors X_m^* increases due to transitions from X_m to X_m^* at a ligand-dependent rate k(l). The number X_m^* decreases due to the de-methylating action of CheB and due to transitions to the inactive state X_m at a ligand-dependent rate k'(l). The dynamics of $X_m^*=A$ therefore is given by the sum over the rates of all of these transitions with appropriate signs

(7.3.5) $dA/dt = k(l) X_m - k'(l) A - V_B B A$

We want to rearrange this equation so that the first term is proportional to $X_{m,t}=X_m+A$ (analogous to the heater power P in exercise 7.2, Eq. 7.2.1). For this purpose, we add and subtract $k(l) A$, to find

(7.3.6) $dA/dt = k(l) X_{m,t} - (k'(l)+k(l)) A - V_B B A$

Thus, we end up with an integral feedback system, analogous to Eq 7.2.1, 7.2.3, in which

(7.3.7) $dA/dt = a X_{m,t} - b A$

(7.3.8) $d X_{m,t} / dt = -K (A-A_{st})$

where $a=k(l)$, $b=-(k'(l)+k(l))-V_B B$ and $K=V_R R$.

To restate the analogy, think of A as the temperature and $X_{m,t}$ as the power to the heater in problem 7.2. As shown in problem 7.2, the steady-state activity A_{st} does not depend on any of the parameters *a, b* or *K*, and in particular on the ligand level that enters only through *k(l)* and *k'(l)* in the parameters *a* and *b*. Thus A_{st} does not depend on the level of attractant ligand (or repellent ligand), and exact adaptation is achieved.

Exercises, chapter 8

8.1 *Diffusion from both sides:* A morphogen is produced at both boundaries of a region of cells that ranges from x=0 to x=L. The morphogen diffuses into the region and is degraded at rate α . What is the steady state concentration of the morphogen as a function of position? Assume that the concentration at the boundaries is $M(0)=M(L)=M_o$. Under what conditions is the concentration of morphogen at the center of the region very small compared to M_o?

Hint: The morphogen concentration obeys the diffusion-degradation equation at steady state:

$D\, d^2 M / d x^2 - \alpha M = 0$

The solutions of this equation are

$M(x)= A \exp(-x/\lambda)+B \exp(x/\lambda)$

Find λ, A and B that satisfy the diffusion degradation equation and the boundary conditions.

Solution:

Differentiating the proposed solution once results in :

(8.1.1) $$\frac{dM}{dx} = -\frac{A}{\lambda}e^{-x/\lambda} + \frac{B}{\lambda}e^{x/\lambda}$$

Differentiating once again:

(8.1.2) $$\frac{d^2M}{dx^2} = \frac{A}{\lambda^2}e^{-x/\lambda} + \frac{B}{\lambda^2}e^{x/\lambda}$$

Substituting this into the steady state equation :

(8.1.3) $$D\frac{d^2M}{dx^2} - \alpha M = \frac{AD}{\lambda^2}e^{-x/\lambda} + \frac{BD}{\lambda^2}e^{x/\lambda} - \alpha A e^{-x/\lambda} - \alpha B e^{x/\lambda} = 0$$

By equating to zero the coefficients of $\exp(x/\lambda)$ (or $\exp(-x/\lambda)$) we get :

(8.1.4) $$\frac{AD}{\lambda^2} - \alpha A = 0 \Rightarrow \lambda = \sqrt{\frac{D}{\alpha}}$$

To find A and B we shall use the boundary conditions:

(8.1.5) $$M(0) = A + B = M_0$$

(8.1.6) $$M(L) = Ae^{-L/\lambda} + Be^{L/\lambda} = M_0$$

Solving these two equations for A and B results in the following solution:

(8.1.7) $$A = M_0 \frac{e^{L/\lambda} - 1}{e^{L/\lambda} - e^{-L/\lambda}}$$

(8.1.8) $$B = M_0 \frac{1 - e^{-L/\lambda}}{e^{L/\lambda} - e^{-L/\lambda}}$$

8.2 Diffusion with degradation at boundary: A morphogen is produced at x=0 and enters a region of cells where it is not degraded. The morphogen is, however, strongly degraded at the other end of the region, at x=L, such that every molecule of M that reaches x=L is immediately degraded. The boundary conditions are thus M(0)=Mo, and M(L)=0.

(a) What is the steady state concentration profile of M?

(b) Is patterning by this mechanism robust to changes of the concentration at the source, $M(0)=M_o$?

Hint: The morphogen obeys a simple equation at steady state:

$D\, d^2 M / d x^2 = 0$

Try solutions of the form

$M(x) = A x + B$, and find A and B such that $M(x=L)=0$ and $M(x=0)=M_o$.

Next, find the position where M(x) equals a threshold T, and find the changes in this position upon a change of Mo.

Solution:

By solving the coefficients A and B of the suggested linear form of the steady state morphogen concentration we get :

(8.2.1) $M(x)=-(M_0/L)*x+M_0$

Let us calculate the position at which the morphogen level are at a certain threshold level T :

(8.2.2) $T=-(M_0/L)*x_0+M_0 \Longrightarrow x_0=L(1-T/M_0)$

If the morphogen level at the source changes to $M_0'=aM_0$ the change in position is :

(8.2.3) $\delta x_0=x_0'-x_0=LT(1/M_0'-1/M_0)=LT(1-a)/(aM_0)$

The change in position δx_0 implicitly depends on the position x_0 through T (8.2.2). For low thresholds, which are related to positions close to the right boundary x=L, the effect of morphogen changes at the source are less than for high thresholds, which are related to positions close to the source, x=0.

8.3 *Polynomial self-enhanced degradation*: Find the steady-state concentration profile of a morphogen produced at x=0. The morphogen diffuses into a field of cells, with nonlinear self-enhanced degradation described by

$\partial M / \partial t = D \partial^2 M / \partial x^2 - \alpha M^n$

When is patterning with this profile robust to the level of M at the boundary, M_0?

Hint: try a solution of the form $M(x) = a (x+b)^m$ and find the parameters a and b in terms of D, M_0 and α.

Solution:

Differentiating the proposed solution once and twice leads to the following terms:

(8.3.1) $$\frac{dM}{dt} = ma(x+b)^{m-1}$$

(8.3.2) $$\frac{d^2M}{dt^2} = m(m-1)a(x+b)^{m-2}$$

Inserting into the steady state equation:

(8.3.3) $$D\frac{d^2M}{dt^2} - \alpha M^n = Dam(m-1)(x+b)^{m-2} - \alpha a^n (x+b)^{mn} = 0$$

Equation (8.3.3) can hold for any x only if the exponents of the two terms (x+b) are equal and the coefficient is equal to zero :

(8.3.4) $$nm = m-2 \Rightarrow m = \frac{2}{1-n}$$

(8.3.5) $$aD\frac{2}{(1-n)}\frac{(1+n)}{(1-n)} - \alpha a^n = 0$$

Solving for a we get :

(8.3.6) $$a = \left(\frac{D}{\alpha}\frac{(2n+2)}{(1-n)^2}\right)^{\frac{1}{n-1}}$$

To find b we shall use the boundary condition at x=0 :

(8.3.7) $$M(0) = ab^m = M_0 \Rightarrow b = \left(\frac{M_0}{a}\right)^{(1-n)/2} = M_0^{\frac{1-n}{2}}\left(\frac{D}{\alpha}\frac{2n+2}{(1-n)^2}\right)^{\frac{1}{2}}$$

Using n=2, as in the example presented in the book, we get:

$A=6D/\alpha$, $B=(6D/\alpha M_0)^{1/2}$.

8.4 *Robust timing:* A signaling protein X inhibits pathway Y. At time t=0, X production stops, and its concentrations decays due to degradation. The pathway Y is

activated when X levels drop below a threshold T. The time at which Y is activated is t_Y. Our goal is to make t_Y as robust as possible to the initial level of X, X(t=0)=Xo.

(a) Compare the robustness of t_Y in two mechanisms, linear degradation and self-enhanced degradation (note that in this problem, all concentrations are spatially uniform).

$\partial X / \partial t = - \alpha X$

$\partial X / \partial t = - \alpha X^n$

Which mechanism is more robust to fluctuations in Xo? Explain.

Hint: try solutions of the form $X(t)=a (t+b)^m$

(b)Explain why a robust timing mechanism requires a rapid decay of X at times close to t=0.

Solution:

(a) The solution of the first equation $\partial X / \partial t = - \alpha X$

is a simple exponential form:

(8.4.1) $X(t)=X_0 \exp(-\alpha t)$

To solve the second equation: $\partial X / \partial t = - \alpha X^n$

we shall differentiate the proposed form $X(t)=a (t+b)^m$

and substitute :

(8.4.2) $am(t+b)^{m-1}=-\alpha a^n (t+b)^{mn}$

The solution will hold for all times *t* only if the exponents of (t+b) and the coefficients of the (t+b) terms are equal :

(8.4.3) $mn=m-1 \Longrightarrow m=1/(1-n)$

(8.4.4) $am=a/(1-n)=-\alpha a^n \Longrightarrow a=[\alpha(n-1)]^{1/(1-n)}$

Finally solving at t=0 we get:

(8.4.5) $X(0)=ab^m=ab^{1/(1-n)} \Longrightarrow b=\dfrac{X_0^{1-n}}{\alpha(n-1)}$

The solution is:

(8.4.6) $X(t)=a(t+b)^m= [\alpha(n-1)]^{1/(1-n)}(t+\dfrac{X_0^{1-n}}{\alpha(n-1)})^{1/(1-n)} = (\alpha(n-1)t + X_0^{1-n})^{1/(1-n)}$

To compute the robustness of Y activation to fluctuations in X_0 we shall calculate the time t_Y at which Y is activated. For the exponential form:

(8.4.7) $$T = X_0 e^{-\alpha t_Y} \Rightarrow t_Y = -\frac{1}{\alpha}\log(\frac{T}{X_0})$$

and the time difference for Y activation caused by an initial value of X_0' is :

(8.4.8) $$\delta t_Y = -\frac{1}{\alpha}[\log(\frac{T}{X_0'}) - \log(\frac{T}{X_0})] = -\frac{1}{\alpha}\log(\frac{X_0}{X_0'})$$

For the power-law form, we get we will assume that X_0 is much larger than T. At a late enough time t, this assumption reduces equation (8.4.6) to:

(8.4.9) $$T = [\alpha(n-1)t]^{1/(1-n)} \Rightarrow t_Y = \frac{T^{1-n}}{\alpha(n-1)}$$

The power-law form is more robust to fluctuations, and does not depend on X_0 at all. The assumption of large X_0 is equivalent to a rapid decay at times close to t=0, as the degradation depends non-linearly on X levels.

8.5 Activator accumulation versus repressor decay (harder problem): Compare the robustness of t_Y in problem 8.4 to an alternative system, in which X is an activator, which begins to be produced at t=0, activating Y when it exceeded threshold T. Assume either linear or nonlinear degradation of X. Is the accumulating activator mechanism more or less robust to the production rate of X than the decaying repressor mechanism? (The production rate of X in the latter case is related to the initial level of X, Xo). Answer: an activator mechanism is generally less robust to variations in the production rate of X than the decaying repressor mechanism of problem 8.3.

Solution:

The temporal solution of X(t) for a repressor is :

(8.5.1) $X(t)=X_0\exp(-\alpha t)=(\beta/\alpha)\exp(-\alpha t)$

where we assume that the initial X_0 level is a steady state level of β/α. From (8.4.8) we get that the difference in Y activation caused by fluctuations in β (or X_0) is :

(8.5.2) $$\delta t_Y = -\frac{1}{\alpha}\log(\frac{X_0}{X_0'}) = -\frac{1}{\alpha}\log(\frac{\beta}{\beta'})$$

For an activator :

(8.5.3) $X(t)=(\beta/\alpha)(1-\exp(-\alpha t))$

Assuming that the threshold for activation of the downstream target Y is much smaller than the steady state level (this design is more robust to noise, see Exercise 5.1). In this regime we can use the linear approximation :

(8.5.4) $X(t)=\beta t$

solving for t_Y, the time when X reaches the threshold T :

(8.5.5) $\beta t_Y=T \Longrightarrow t_Y=T/\beta$

to study the robustness to fluctuations in β we note that :

(8.5.6) $\delta t_Y=T(1/\beta'-1/\beta)$

A repressor is thus more robust to changes in production rate β, as the change appears inside a logarithm.

9.1 *At any rate.* Determine the error in the proofreading process in Eq 9.2.9. What conditions (inequalities) on the rates allow for effective kinetic proofreading?

Solution:

The rate of change of [cC] is governed by the rate of collisions of c and C with on-rate k', their dissociation with off-rate k_c, and the rate of formation of [cC*] at rate m:

(9.1.1) $d[cC]/dt= k'c\ C - (m+k_c)\ [cC]$

so that at steady-state, in which d[cC]/dt=0, we have

(9.1.2) $[cC]=k'\ c\ C/(m+k_c')$

Similarly, for [cC*], that is produced at rate m, dissociates at rate l_c and produces a product with rate v:

(9.1.3) $d[cC^*]/dt= m[cC]-(v+l_c)\ [cC^*]$

So that at steady-state, using Eq 9.1.3, we have

(9.1.4) $[cC^*]=m/(v+l_c)\ [cC]= m\ k'\ c\ C/(v+l_c)\ (m+k_c)$

Similar considerations for the wrong ligand d can be made, noting that for d the on-rate k', the complex formation rate m and the product formation rate v are the same as for c, but that the off rates k_d and l_d are larger than the corresponding rates for c due to the weaker affinity of d to C. Thus

(9.1.5) $[dC^*]= m\ k'\ d\ C/(v+l_d)\ (m+k_d)$

The error rate is the ratio of wrong and correct production rates $v[dC^*]/v[cC^*]$

(9.1.6) $f= v[dC^*]/v[cC^*] = d\ (v+l_c)\ (m+k_c)\ /\ c\ (v+l_d)\ (m+k_d)$

When $v<<l_c$ and l_d, and when $m<< k_c$ and k_d , we have the minimal error rate in this process

(9.1.7) $f= d\ l_c\ k_c\ /\ l_d\ k_d$

Thus minimal errors require that the complexes [dC] dissociate much faster than the rate of formation of [dC*], and that [dC*] dissociate much faster than the rate of product formation. This gives many opportunities for the wrong ligand to fall off of the complex, before an irreversible step takes place.

In processes where the dissociation from the state C and C* are based on the same molecular site (e.g. the tRNA-codon interaction), we have $l_c=k_c$ and the same for d, so that

(9.1.8) $f= d\ (k_c\ /k_d)^2=f_o^2$

Where f_o is the equilibrium error rate.

9.2 *Detailed balance* Determine the error rate in a proofreading scheme in which transition from [cC] to [c*C] occur at a forward rate m_c, backward rate m_c', transitions [c*C] to c+C occur at forward rate l_c and back rate l_c', and corresponding constants for d, and where the product formation rate v is negligible compared to the other rates. Consider the case where all reactions occur at equilibrium. Use the **detailed balance** conditions where the flux of each reaction is exactly equal to the flux of the reverse reaction, resulting in zero net flux along any cycle (better known in biochemistry as the **thermodynamic box** conditions).

(a) Show that detailed balance requires that $k_c\, m_c'\, l_c' = k_c' m_c l_c$,

and the same for d.

(b) Calculate the resulting error rate f. Explain.

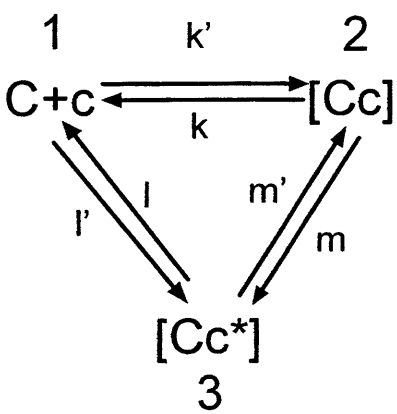
Solution:

(a) The reaction scheme is as follows:

$$\begin{array}{ccccc} & k_c' & m & v & \\ C+c & \Leftrightarrow [Cc] & \Leftrightarrow & [Cc^*] & \longrightarrow \text{product} \\ & k_c & m' & l_c \Updownarrow l_c' & \\ & & & C+c & \end{array}$$

Where k' is the rate of formation of [Cc] from C+c and l' denote the rate of formation of [Cc*] from C+c.

The diagram for the overall reaction is:



by defining K=k'/k, L=l'/l, M=m'/m we can write at steady state:

$K=\exp(-\Delta G_{21}/RT)$

$L=\exp(-\Delta G_{31}/RT)$

$M=\exp(-\Delta G_{23}/RT)$

These equations hold only for chemical species at equilibrium.

From conservation of energy we now have:

$\Delta G_{21}+\Delta G_{32}=\Delta G_{31}$

or equivalently:

$-RT\ln(K)+RT\ln(M)=-RT\ln(L)$

which results in:

K/M=L, or k'm/km'=l'/l

(b) The detailed balance equations (neglecting v) are:

(9.2.1) $k_c'Cc+m'[Cc^*]=(k_c+m)[Cc]$

(9.2.2) $m[Cc]+l_c'Cc=(l_c+m')[Cc^*]$

The solutions to these two equations are:

(9.2.3) $$[Cc]=\frac{k_c'l_c+(k_c'+l_c')m_c'}{m_c'k_c+(m_c+k_c)l_c}$$

(9.2.4) $$[Cc^*]=\frac{m_ck_c'+(k_c+m_c)l_c'}{m_c'k_c+(m_c+k_c)l_c}$$

Assuming all specificity is in the off-rates, $m_c=m_d=m$, $m_c'=m_d'=m'$, $k_c'=k_d'=k'$, $l_c'=l_d'=l'$, the error rate f is :

(9.2.5)

$$f=\frac{v[Cd^*]}{v[Cc^*]}=\frac{m_dk_d'+(k_d+m_d)l_d'}{m_d'k_d+(m_d+k_d)l_d}*\frac{m_c'k_c+(m_c+k_c)l_c}{m_ck_c'+(k_c+m_c)l_c'}=\frac{[mk'+(k_d+m)l']*[m'k_c+(m+k_c)l_c]}{[m'k_d+(m+k_d)l_d]*[mk'+(k_c+m)l']}$$

Let's examine how certain assumptions on this general error rate lead to the one seen in (9.2.14). First we assume that the intermediate states [Cc*] ([Cd*]) are energy driven, so that that the reaction rates $l_c'=l_d'=0$, as the rate of spontaneously forming these high energy states is negligible. In addition, we can keep the rate m' very low. An additional assumption we shall make is that $m<<k_c$ and $m<<k_d$. Using these assumptions, and making the appropriate approximations (such as $k_d+m\sim k_d$) we get:

(9.2.6) $$f=\frac{k_cl_c}{k_dl_d}$$

as in 9.1.

9.3 *Optimal tRNA concentrations:* In order to translate a codon, different tRNAs randomly enter the ribosome, attempt to bind the codon, and are ejected if they do not match. That means that, on average, many different tRNAs need to be sampled for each codon until the correct match is found. Still, the ribosome manages to translate several dozen codons per second. We will try to consider the optimal relations between the concentrations of the different tRNAs, which allow the fastest translation process.

(a) Let the concentration of tRNA number j (j goes from one to 64, the number of different types of tRNAs in the cell) be c_j. The relative concentration of tRNA number j is $r_j = c_j / \Sigma c_j$. Suppose that each tRNA spends an average time t_0 in the ribosome before it is ejected or used for translation. What is the average time needed to find the correct tRNA for codon j? Assume that there is no delay between ejection of a tRNA and the entry of a new tRNA, and neglect the ejection of a correct tRNA.

(b) Suppose that the probability of codon j in the coding region of genes in the genome is p_j. What is the optimal relative concentration of each tRNA that allows the fastest translation? Use a Lagrange multiplier to make sure that $\Sigma r_j = 1$.

Solution :

(a) When a tRNA enters the ribosome, the probability that it is the correct codon is r_j. Thus, on average one must try $1/r_j$ tRNAs before the correct one enters the ribosome. Hence, the average time to find the correct tRNA for codon j is

$T_j = t_0 / r_j$.

(b) The time to translate the average codon is the sum of the times T_j weighted by the codon probabilities in the genome:

$T=\Sigma T_j p_j = \Sigma p_j t_0 / r_j$

To minimize the translation time, we need to minimize T. Taking the derivative of T with respect to each r_j, we look for the relative concentrations that yield a minimum and thus have zero derivative, using a LaGrange multiplier L to make sure that $\Sigma r_j = 1$

$d T / d r_j = d/dr_j (\Sigma p_j t_0 / r_j + L \Sigma r_j) = - t_0 p_j / r_j^2 + L = 0$

Solving for r_j, and using a value of L such that $\Sigma r_j = 1$, yields an optimal r_j that is related to the square root of the codon probability p_j:

$r_j^{opt} = \sqrt{p_j} / \sum_j \sqrt{p_j}$

Thus, the rarer the codon, the lower the concentration of its tRNA. The relation between the abundance of a certain tRNA and the abundance of its codon in the DNA, p_i (also known as codon usage) follows a square-root law. If an organism's DNA contains 100 times as many i codons as j codons, it will produce 10 times as many corresponding $tRNA_i$ as $tRNA_j$.

9.4 *Optimal genetic code for minimizing errors:* In this exercise we consider an additional mechanism for reducing translation errors, based on the structure of the genetic code. For simplicity, we will consider a code based on an alphabet of two letters (0 and 1), and where codons have two letters each. Thus, there are four codons ([00],[01],[10] and [11]). This genetic code encodes two amino acids, A and B (and no stop codons). Each amino acid is assigned two of the four codons.

(a) What are the different possible genetic codes?

(b) Assume that mis-reading errors occur, such that a codon can be mis-read as a codon that differs by one letter (e.g. [00] can be misread as [01] or [10], but not as [11]). Which of the possible codes make the fewest translation errors, incorporating the wrong amino acid?

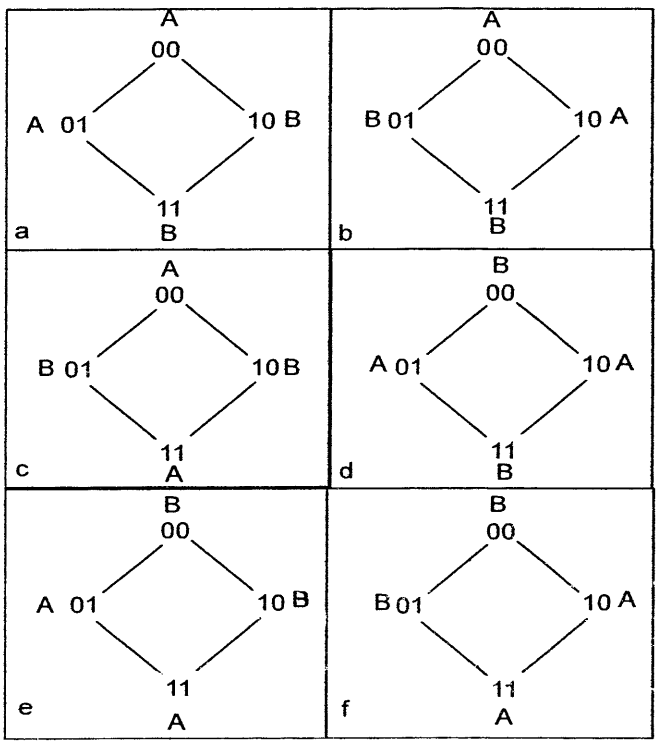
(c) Assume that the first letter in the codon is mis-read at a higher probability then the second letter (e.g. [00] is misread as [10] more often than as [01]). Which of the codes have the lowest translation errors?

(d) Study the real genetic code in Fig 9.1b. Compare the grouping of codons that correspond to the same amino acid. How can this ordering help reduce translation errors? Based on the structure of the genetic code, can you guess which positions in the codon are most prone to mis-reading errors? Can you see in the code a reflection of the fact that U and C in the third letter of the codon can not be distinguished by the translation machinery ('third base wobble')?

(e) In the real genetic code, chemically similar amino acids tend to be encoded by similar codons. Discuss how this might affect the impact of translation errors on the organisms' fitness.

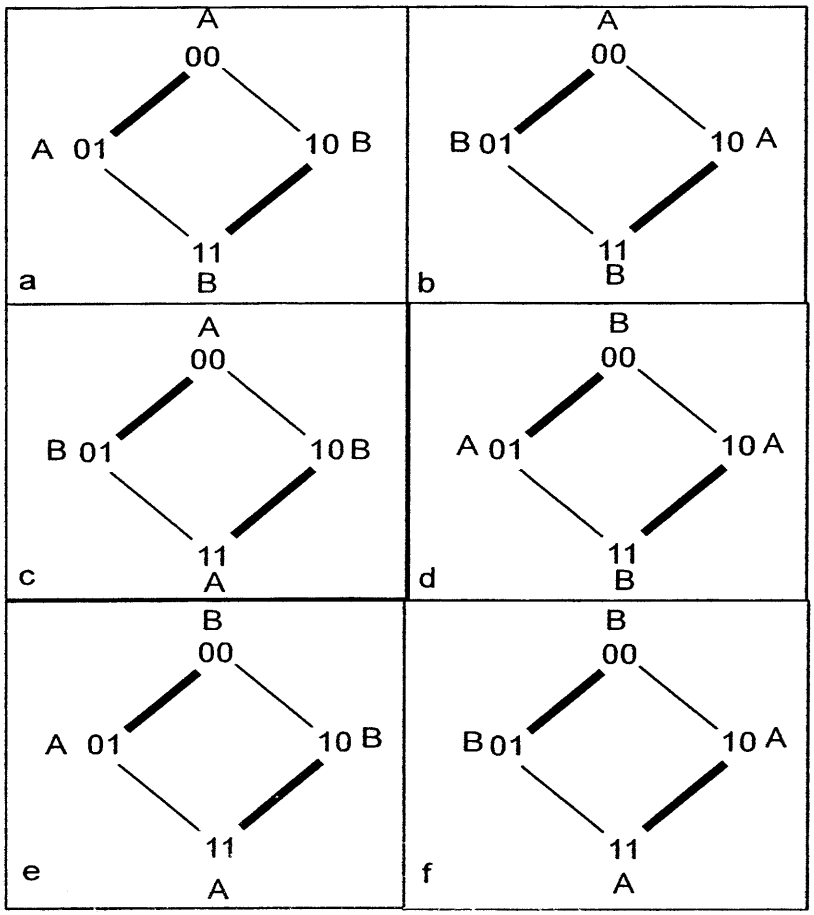
Solution :

(a) There are 6 possible genetic codes in which each amino acid is assigned 2 codons. These are depicted as labeled graphs where codons are nodes and edges connect two codons differing by one letter. The labels are the assigned amino acids (A/B).



(b) Codes a,b,e,f are all optimal in terms of robustness to translational errors. In all four codes there are only two misread events that lead to an incorporation of the wrong amino acid. The four codes are equivalent in terms of the overall error-load.

(c) Now the graphs are weighted graphs, where bold edges connect pairs of codons that are misread with a higher probability :



The degeneracy of the optimal code in (b) is now removed. Codes a and f are more optimal now than codes b and e. This is because assigning code-words which have a higher probability to be misread to the same amino-acid will on average reduce the translation errors.

(d) In the universal genetic code pairs of codons which differ by one letter are either assigned to the same amino acid or to chemically related amino-acids. The most error-prone position is the third base, reflected in the fact that most of the codons differing by only a third base encode the same amino-acid. Next is the first base, reflected by codons of Leu for example where CUU and UUG differ in the first letter and are assigned to the same amino-acid. The least error-prone position in the code is the second letter, which is read with the highest fidelity. This is reflected by the fact that no pair of codons differing in only the second letter are ever assigned to the same amino-acid. We can therefore assume that misread events between such codon pairs are rare.

The wobble constraint on the translation mechanism states that U and C cannot be distinguished in the third codon position. This is reflected in the code by the fact that all pairs of codons which differ have either U/C in the third position are always assigned to the same amino-acid. If a pair of codons, say UGU and UGC were assigned to different amino-acids, 50% of the time the resulting amino-acid would not be the one intended for. In the toy-model code of (c) the 2 position 2 letter code had a 'wobble' in the first letter (the probability of misreading 00 as 10 was higher than misreading 00 as 01).

(e) In the real code there are 64 codons and 20 amino acids. The assignment of the same amino acid to codon pairs which have a high probability to be misread minimizes the impact of translation errors. An ideal situation for the minimization of translation errors was one in which all codons encoded a single amino acid. However, the need for diversity of protein structure requires more than one amino acid, and indeed we see 20 different amino acids, some of which are more similar to each other (e.g. Glu and Asp are both polar, Leu and Ile are both non-polar). In this code a further reduction of the impact of errors can be achieved by assigning codon pairs which have a high probability of being misread to either the same amino acid (as in the wobble case of (d)) or to chemically related amino acids (note for example the positions of the codons for Asp and Glu).

10.1 *Limiting enzyme:* Protein X is an enzyme that acts on a substrate to provide benefit to the cell. L is the substrate concentration. Calculate the fitness function *f(X,L)* assuming a cost that is linear in the protein concentration , c~~ηX, and benefit that is a Michaelis-Menten function of the protein concentration , b(L,X) = b_0 L X/(X+K), appropriate for cases where enzyme X, rather than its substrate, is limiting. Calculate the optimal enzyme level as a function of L and K.

Solution:

The fitness function is the benefit obtained from the ligand minus the cost of enzyme production:

$f_L(X) = b - c = b_0 LX/(X+K) - \eta X$

The maximal fitness can be obtained by differentiating the fitness and setting to 0:

$d f_L / d X = 0$

yielding:

$$X_{opt} = \sqrt{\frac{b_0 LK}{\eta}} - K$$

That is the more lignad L, the higher the enzyme level is.

10.2 For exercise 10.1, what is the minimal substrate level L_c required for maintenance of the gene for X by the organism? When is the gene that encodes X lost? Explain.

Solution:

The minimal substrate level L_c required for X gene maintainance can be found by setting the optimal X expression level to 0:

$$X_{opt} = \sqrt{\frac{b_0 L_c K}{\eta}} - K = 0 \Rightarrow L_c = \frac{\eta K}{b_0}$$

This result makes sense, as the more benefit obtained from a single ligand ,b_0, the lower ligand levels for which the overall benefit is higher than the cost of protein X production. Conversely, the higher the cost per protein unit η is the higher the minimal ligand level L_c for which it pays off to produce enzyme X.

10.3 *Optimal expression of a subunit.*

(a) Multiple units of protein X act together in a multi-subunit complex. The benefit is a Hill function $b(X) = b_0 X^n / (K^n+X^n)$, and the cost function is linear in X. What is the optimal protein level? Explain.

(b) Protein X brings benefit to the cell only when its concentration exceeds Xo, so that $b(X) = \theta(X>Xo)$, where θ is the step function. What is the optimal expression level of X?

Solution:

(a) The fitness function is the benefit minus the cost:

$$f_L(X) = b - c = \frac{b_0 X^n}{X^n + K^n} - \eta X \tag{10.3.1}$$

Differentiating and equating to zero:

$$\frac{df_L(X)}{dX} = b_0 \frac{nX^{n-1}(X^n + K^n) - nX^{2n-1}}{(X^n + K^n)^2} - \eta = 0 \tag{10.3.2}$$

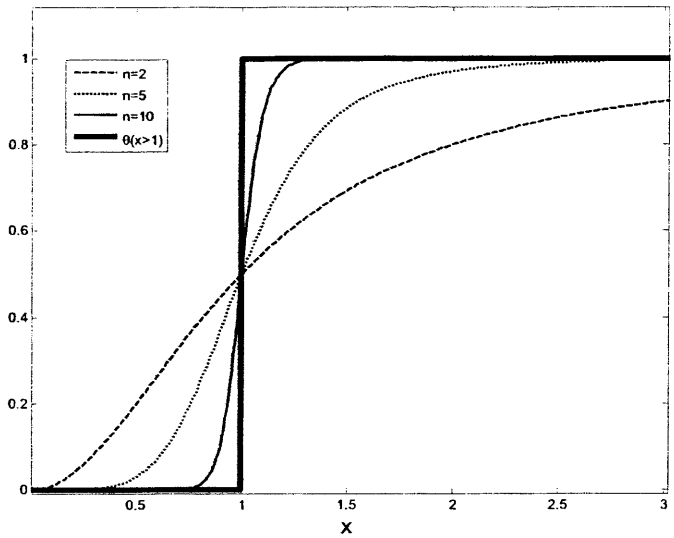
$$b_0 nX^{n-1}K^n = \eta(X^n + K^n)^2 \tag{10.3.3}$$

If we substitute n=1 we get the solution of problem 10.2. The equation can be numerically solved to find $X_{opt.}$

(b) For very large n the benefit of the fitness function in (a) approaches a step function:

$$\frac{X^n}{X^n + K^n} \xrightarrow{n\to\infty} \theta(X - K)$$

as can be seen in the following figure:



For this benefit function the optimal X level will be X_{opt}=K. if X<K there is no benefit and thus it makes no sense to produce the protein. When X>K the benefit stays constant (a step function) but the cost increases, therefore the minimal cost will be the lowest X, X=K.

10.4 *Cost function:*

(a) Derive the cost function in Eq 10.2.4, based on a limiting resource R, such that the growth rate is equal to $f = f_0 R / (K_R + R)$. Each unit of protein Z reduces R by a small amount ε.

(b) In bacterial cells, the resource R often increases as the growth rate decreases. For example, the fraction of free ribosomes increases as growth rate slows, because at high growth rates the ribosomes are mostly engaged in making new ribosomes. This effect can be added to the model to find a similar cost function at the low to intermediate expression levels of Z relevant to the experiments described in this chapter, but with no divergence at high Z. Assume that $R = m / f$, where f is the growth rate and m is a parameter. Derive the cost function in this case.

Solution:

(a) The burden of Z production can be described as a reduction in the internal resource R, such that that each unit of protein Z reduces the resource by a small amount ε, so that R goes to $R - \varepsilon Z$. Hence the cost is as in Eq 10.2.4:

(10.4.1)

$$C = \frac{f(0)-f(Z)}{f(0)} = 1 - \frac{f(Z)}{f(0)} = 1 - \frac{(R-\varepsilon Z)(K+R)}{R(K+R-\varepsilon Z)} = \frac{\varepsilon Z K}{R(K+R-\varepsilon Z)} = \frac{\eta Z}{1 - Z/M}$$

where the initial reduction per subunit of Z is $\eta = K\varepsilon / R\,(K+R)$ and the parameter M is $M = (K+R)/\varepsilon$. Note that the cost can never diverge, because when Z depletes all of the resource R, that is when $Z=R/\varepsilon$, one finds $f(Z)=0$ and the cost is equal to $f(0)$.

(b) The growth rate obeys the following self consistent equation (for simplicity we'll assume the constant $f_0=1$):

(10.4.2)
$$f(Z) = \frac{R-\varepsilon Z}{K+R-\varepsilon Z} = \frac{\frac{m}{f} - \varepsilon Z}{K + \frac{m}{f} - \varepsilon Z} = \frac{m - \varepsilon Z f}{(K-\varepsilon Z) f + m}$$

The solution for f(Z) is :

(10.4.3)
$$f(Z) = \frac{-(m+\varepsilon Z) + \sqrt{(m+\varepsilon Z)^2 + 4m(K-\varepsilon Z)}}{2(K-\varepsilon Z)}$$

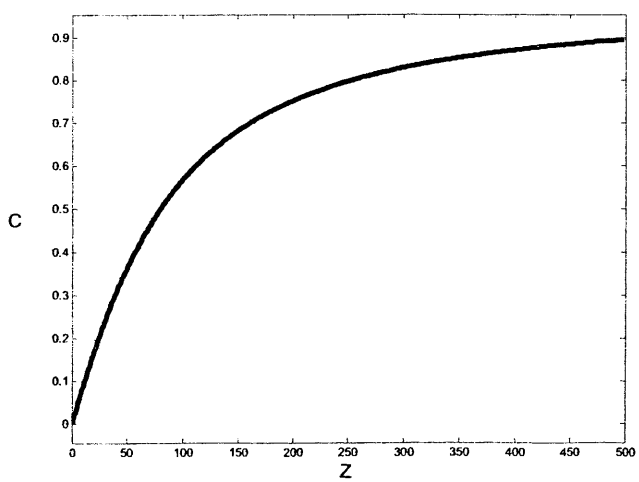
and the relative cost function is :

(10.4.4)

$$C = \frac{f(0)-f(Z)}{f(0)} = 1 - \frac{-(m+\varepsilon Z) + \sqrt{(m+\varepsilon Z)^2 + 4m(K-\varepsilon Z)}}{2(K-\varepsilon Z)} * \frac{2K}{-m + \sqrt{m^2 + 4mK}}$$

This function does not diverge with Z.

Plotting the cost function for m=1, K=100, ϵ=0.2:



11.1 *Optimization versus historical accident:* Imagine a population of organisms with a regulatory mechanism in place for a certain gene. Conditions change, and the opposite mode of regulation is now more optimal for that gene, in the sense that it has a lower error load. The demand for the gene in the new environment is p, and the error loads associated with errors in the high and low expression states are Δf_1 and Δf_0. Mutants with the opposite mode arise in the population, but they can only fully replace the original population if their fitness advantage exceeds a minimal value s_{min}

(a) Calculate the conditions on the demand p in which the mutants with the optimal mode can take over the population.

(b) When is the mode of regulation determined by historical precedent? Explain.

Solution:

(a) The error load of a repressor is:

(11.1.1) $$E_R = p\,\Delta f_1$$

From Eq 11.2, the error load of an activator is:

(11.1.2) $$E_A = (1-p)\,\Delta f_0$$

The fitness is minus the error load, hence assuming an existing repressor solution, an activator solution will fix in the population if the fitness difference exceeds the minimial selection value s_{min}:

(11.1.3) $$(-E_A)-(-E_R) > s_{min}$$

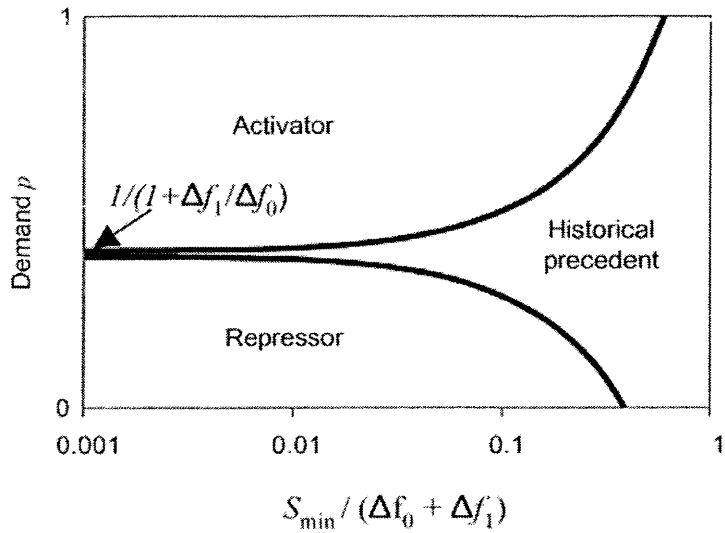
$$\Rightarrow p > \frac{1}{1+\Delta f_1/\Delta f_0} + \frac{s_{\min}}{\Delta f_0 + \Delta f_1}$$

In the case of an existing activator solution, a repressor will fix in the population if:

(11.1.4) $$(-E_R)-(-E_A) > s_{min}$$

$$\Rightarrow p < \frac{1}{1+\Delta f_1/\Delta f_0} - \frac{s_{\min}}{\Delta f_0 + \Delta f_1}$$

These regimes can be seen in the following figure:



(b) When the demand values are in the range

$$\frac{1}{1+\Delta f_1/\Delta f_0} - \frac{s_{min}}{\Delta f_0 + \Delta f_1} < p < \frac{1}{1+\Delta f_1/\Delta f_0} + \frac{s_{min}}{\Delta f_0 + \Delta f_1} \quad (11.1.5)$$

The fitness advantage of the activator/repressor solution relative to the alternative solution is not enough to fix the mutation in the population. The original solution will prevail. This regime can be defined as an historical precedent.

11.2 *Error-load of noise in protein expression*: The expression of proteins varies from cell to cell. This means that different cells deviate from the optimal expression level. In this exercise we will calculate the average reduction in fitness due to such variations, for the case of the *lac* system. The fully induced *lac* promoter has a cell-cell variation in expression with coefficient of variation (standard deviation of protein level Z divided by the mean) of about CV=0.1 in the fully induced state (Elowitz, 2002). The relative fitness function for this exercise, similar to the function we saw in chapter 10, is $f(Z)= -\eta\, Z/(1-Z/M) +\delta\, Z$, with $\eta \sim 0.05 Z_0^{-1}$ and $M \sim 2Z_0$, where Z_0 is the fully induced expression level.

(a) Show that the mean reduction in fitness due to small cell-cell variations in Z is $\Delta f = C \langle \Delta Z^2 \rangle$, where C is the curvature of the fitness function near its maximum $C=1/2\, d^2f/dZ^2$ and the brackets $\langle\rangle$ denote a population average. Hint: use a Taylor expansion of f(Z) near its maximum $Z=Z_0$.

(b) Compute the mean reduction in fitness due to variations in Z in the fully induced state (that is when the average value of Z in the population is Zo).

Solution:

(a) The Taylor expansion of the fitness function around its maximum point $Z=Z_0$ is:

(11.2.1)

$$f(Z_0+\Delta Z)=f(Z_0)+\left.\frac{\partial f}{\partial Z}\right|_{Z=Z_0}\Delta Z+\left.\frac{1}{2}\frac{\partial^2 f}{\partial Z^2}\right|_{Z=Z_0}(\Delta Z)^2=f(Z_0)+\left.\frac{1}{2}\frac{\partial^2 f}{\partial Z^2}\right|_{Z=Z_0}(\Delta Z)^2$$

as the first derivative of f is 0 at the maximum point Z_0. Thus the mean reduction in fitness is:

(11.2.2) $$<\Delta f>=f(Z_0+\Delta Z)-f(Z_0)=\left.\frac{1}{2}\frac{\partial^2 f}{\partial Z^2}\right|_{Z=Z_0}<(\Delta Z)^2>=C<(\Delta Z)^2>$$

(b) The fitness function is:

(11.2.3) $$f(Z)=-\frac{\eta Z}{1-Z/M}+\delta Z=-\frac{0.05Z/Z_0}{1-Z/2Z_0}+\delta Z=\delta Z-\frac{0.1Z}{2Z_0-Z}$$

Differentiating f twice gives:

(11.2.4) $$\left.\frac{\partial^2 f}{\partial Z^2}\right|_{Z=Z_0}=\left.\frac{2}{5}\frac{Z_0}{(Z-2Z_0)^3}\right|_{Z=Z_0}=-\frac{2}{5Z_0^2}$$

The coefficient of variance is the standard deviation divided by the mean:

(11.2.5) $$CV(Z)=\frac{\sqrt{<(Z-Z_0)^2>}}{Z_0}=\frac{\sqrt{<(\Delta Z)^2>}}{Z_0}\Rightarrow<(\Delta Z)^2>=CV^2Z_0^2=0.01Z_0^2$$

Finally, inserting (11.2.4) and (11.2.5) into (11.2.2):

(11.2.6) $$<\Delta f>=C<(\Delta Z)^2>=\frac{1}{2}\frac{-2}{5Z_0^2}0.01Z_0^2=0.002\sim 0.2\%$$

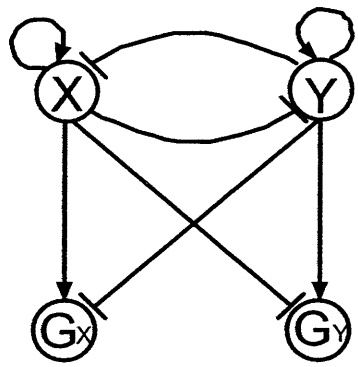
11.3 *Demand rules for developmental genes:* Consider a cell which, during the developmental process of the organism, can assume either fate A or fate B. A set of genes G_A is expressed in fate A and not in fate B, and a different set G_B is expressed in fate B and not in A, This cell-fate decision is regulated by two transcription-factors X and Y. X activates its own transcription and represses the transcription of Y, whereas Y activates its own transcription and represses the transcription of X. Furthermore, X transcriptionally activates G_A and represses G_B, and Y has the opposite effect, activating G_B and repressing G_A.

(a) Draw the transcription network in this case.

(b) Explain the mode of regulation of each gene in terms of the demand rules.

Solution:

(a)



(b) The double negative feedback loop composed of X-Y can result in two steady states – X high, Y low and X low, Y high (chapter 6.2). In cells which express the G_X genes X will be high and Y low. G_X genes are at high demand in these cells whereas G_Y genes are in low demand. Indeed G_X genes are activated by the active transcription factor X, whereas G_Y genes are repressed by X. The same logic implies to cells in which the Y transcription factor is high. These cells express G_Y genes in high demand and G_X in low demand, reflected in the mode of regulation of Y which is an activator for G_Y and repressor for G_X.

11.4 *Error-load of two-input systems:* Compute the error-loads of all possible four regulatory mechanisms for the *lac* system, according to Table 11.4. Compute the conditions (the range of p_{00} and p_{01}) in which each of the four mechanisms is optimal. Compare your results to Fig 11.6.

Solution:

Based on Table 11.4, the error load for each design is:

(11.4.1) $E_{AA}=p_{00}\Delta f_2'+(1- p_{00}- p_{01})(\Delta f_1+\Delta f_1')$

(11.4.2) $E_{AR}=p_{01}\Delta f_4'+(1- p_{00}- p_{01})\Delta f_1$

(11.4.3) $E_{RA}=p_{00}(\Delta f_2+\Delta f_2')+ p_{01}\Delta f_4+(1- p_{00}- p_{01})\Delta f_1'$ (*)

(11.4.4) $E_{RR}=p_{00}\Delta f_2+p_{01}(\Delta f_4+\Delta f_4')$

* note a typo in Table 11:4 - $\Delta f_1'$ instead of $\Delta f_4'$

Using these formulas we can determine the regimes in which different solutions are optimal. An activator-activator solution is more optimal than the activator-repressor solution when:

(11.4.5) $$E_{AA} < E_{AR} \Rightarrow p_{01} > \frac{\Delta f_1'}{\Delta f_1' + \Delta f_4'} + \frac{\Delta f_2' - \Delta f_1'}{\Delta f_1' + \Delta f_4'} p_{00}$$

Similarly, a repressor-repressor solution is optimal when $E_{RR} < E_{AR}$:

(11.4.6) $$E_{RR} < E_{AR} \Rightarrow p_{01} < \frac{\Delta f_1}{\Delta f_1 + \Delta f_4} - \frac{\Delta f_2 - \Delta f_1}{\Delta f_1 + \Delta f_4} p_{00}$$

It can be shown that the repressor-activator solution is never optimal. This is because There is no regime in which $E_{RA} < E_{AR}$ and $E_{RA} < E_{RR}$ Simultaneously. The three regimes in which E_{RR}, E_{AR}, and E_{AA} are optimal can be seen in figure 11.6. Note that the intersect with the p_{01} axis in Eq 11.4.5 is:

(11.4.7) $$\frac{\Delta f_1'}{\Delta f_1' + \Delta f_4'} = \frac{1}{1 + \Delta f_4' / \Delta f_1'}$$

where as the intersect in Eq 11.4.6 is:

(11.4.8) $$\frac{\Delta f_1}{\Delta f_1 + \Delta f_4} = \frac{1}{1 + \Delta f_4 / \Delta f_1}$$

This intersect is smaller:

(11.4.9) $$\frac{1}{1 + \Delta f_4' / \Delta f_1'} > \frac{1}{1 + \Delta f_4 / \Delta f_1}$$

because in the lac system $\Delta f_4 \sim (Z_4 - Z_3)^2 \sim \Delta f_4' = (Z_4 - Z_2)^2$ but $\Delta f_1 = (Z_2 - Z_1)^2 << \Delta f_1' = (Z_3 - Z_1)^2$. The latter inequality can be appreciated by considering the levels of expression in the lac system, sorted from the smallest expression level Z_1 when there is glucose and no lactose up to the highest expression level Z_4 when there is lactose but no glucose:

$Z_1 = 3*10^{-3}$, $Z_2 = 6*10^{-2}$, $Z_3 = 0.13$, $Z_4 = 1$.

11.5 *Error-load of a feedforward loop (FFL)*: Consider a type-1 coherent FFL (see chapter 4). In this gene circuit, activators X and Y activate gene Z, and X also activates Y so that at steady state, Y levels are zero unless X is transcriptionally active. The regulators X and Y respond to input signals S_x and S_y. The promoter of

gene Z is activated in an additive fashion by X and Y, such that at steady-state the expression level is Z=a X*+bY*.

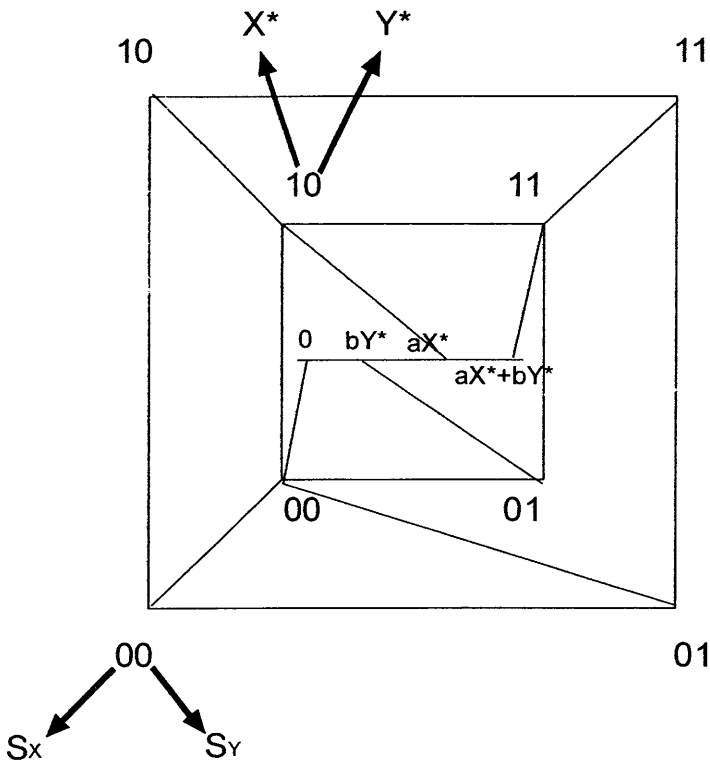
(a) Plot the steady-state relationship between input states (S_x,S_y), internal states (X*,Y*) and output states (Z_1,Z_2,Z_3,Z_4) for all four combinations of S_x,S_y =0 or 1 (Similar to Fig 11.5).

(b) Is there an excluded-state in this case (at steady-state)?

(c) Repeat this for an incoherent type-1 FFL in which X and Y act additively Z=a X*-bY* . Assume that b<a, and that at steady-state, Y levels are zero unless X is transcriptionally active.

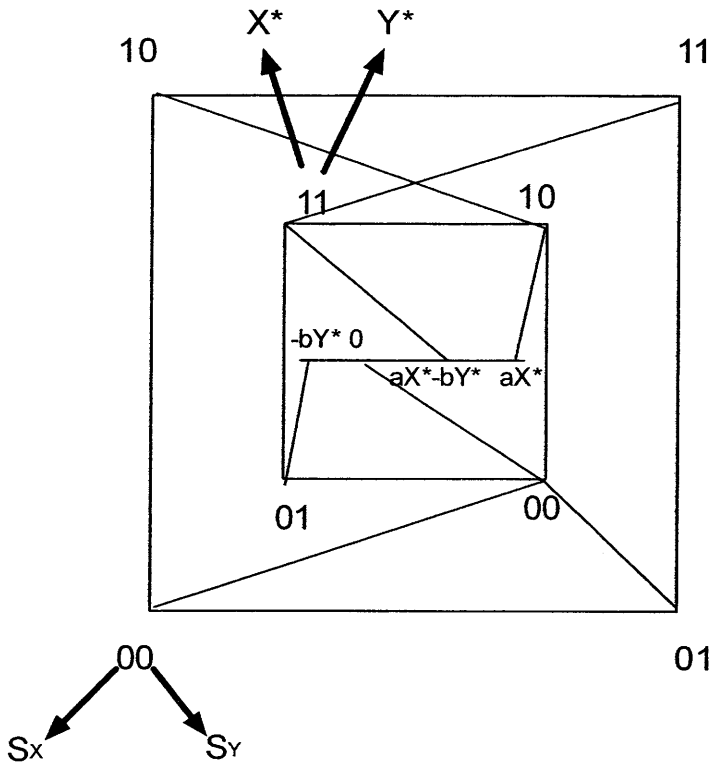
Solution:

(a) For the type I coherent feed-forward loop, the selection diagram is:



(b) The excluded state is X*=0, Y*=1, as X* controls the production of Y*, hence when X* is not present Y* cannot be present.

(c) For the incoherent type I FFL the selection diagram is:



Again the excluded state is X*=0, Y*=1, as X* controls the production of Y*, hence when X* is not present Y* cannot be present.

Final Exam

Introduction to Systems Biology,

Instructor: Uri Alon; *Teaching assistant: Shalev Itzkovitz*

March 27, 2006

Please answer three out of the following four questions.

1. Single input module (SIM): Transcription factor X represses two genes Y_1 and Y_2.

(a) Draw the resulting network, termed a SIM with two target genes.

(b) The repression thresholds for these genes are K_1 and K_2. The repressor X, initially present at steady-state levels due to production at rate β, stops to be produced at time $t=0$, and is degraded/diluted at rate α. The signal S_x is present throughout. What are the times at which the gene products Y_1 and Y_2 reach halfway to their maximal expression?

(c) Design a SIM with a repressor and three target genes in which the genes are activated with equal temporal spacing.

2. Feed-forward loop (FFL) with AND-gate logic. In the C1-FFL, the activator X activates the gene for activator Y, and both X and Y activate output gene Z with AND-gate logic. The activators X and Y have input signals S_x and S_y respectively.

(a) Draw the resulting network.

(b) Provide equations and draw the dynamics of Z, in response to an ON step and an OFF step of S_x. The second input, S_y, is present at all times. Assume Logic input functions throughout.

(c) Are there delays following ON or OFF steps of S_x?

(d) What could be the biological function of such a design (briefly, less than 100 words)?

3.Cost-benefit analysis with limiting enzyme: Protein X is an enzyme that acts on a substrate to provide benefit to the cell. L is the substrate concentration.

a) Calculate the fitness function *f(X,L)* assuming a cost that is quadratic in the protein concentration , $c = \eta X + \eta' X^2$, and benefit that is a Michaelis-Menten function $b(L,X) = b_0 X L / (L+K)$.

(b) Calculate the optimal enzyme level as a function of L and K.

c) Is there a minimal substrate level L_c in which the enzyme is maintained by the organism over evolutionary timescales? Explain briefly. What do you think would happen if the organism lives for a long time in environments with $L<L_c$?

(d) Briefly (in less than 100 words), suggest an experiment to test these results.

4. Perceptrons: Consider a multi-layer perceptron, with two input nodes X_1 and X_2 , two intermediate nodes Y_1 and Y_2, and a single output node Z.

(a) Draw this network, which has 6 edges.

(b) Assume that all nodes are 'integrate and fire': they sum their inputs with weights (each edge has a weight), and produce an output of one only if this sum is greater than a threshold. The output is zero otherwise. Each node can have a threshold of K=1 or K=-1. Weights on the edges can be positive or negative real numbers. The input nodes generate output values between zero and one. Design weights such that this circuit computes the XOR (exclusive-or) function, where Z=1 if either X_1=1 or X_2=1 but Z=0 if both X_1=1 and X_2=1. This function is denoted $Z=X_1$ XOR X_2.

(c) Design weights such that his circuit computes the 'equals' function, in which Z=1 only if X_1 and X_2 are the same (both 0 or both 1), and Z=0 otherwise (that is, $Z=X_1$ EQ X_2).

(d) Draw the output Z in diagrams in which the axes are X_1 and X_2. : To do this, draw the boundaries between regions in which Z=0 and Z=1, for the XOR and EQUAL perceptrons discussed above.

(e) Provide examples of two types of biological systems that appear to use multi-layer perceptrons (answer in less than 100 words).