Master 2 ICFP lectures 2022

Physics of multicellular systems : lecture 1

Vincent Hakim

January 26, 2022

Lecture 1 : introduction to gene networks and modelling.

Insight into the functioning of living cells and organisms is rapidly improving thanks to the recent advances of molecular biology and imaging techniques. However, understanding the precise dynamics, development and evolution of biological organisms remains a major challenge. To make sense of increasing amounts of accurate biological data, one often needs to consider organisms as nonlinear, stochastic interacting systems, the study of which calls for new methods and approaches from theoretical physics. The past decade has seen the development of a new "quantitative biology" that combines experiments and theoretical modelling to explain complex biological phenomena. This set of lectures is an introduction to this young and exploding field, where the concepts and techniques of statistical and nonlinear physics play a primordial role.

1 Some numbers

Species	genome size	protein coding genes	chromosomes
Yeast	13.5 Mb	5800	16
Drosophila	165 Mb	14000	X/Y, 2, 3, 4 (very small)
C elegans	100 Mb	20000	5 autosomes + X
Chicken	1Gb	15000	
Mouse	2.7 Gb	22000 protein coding genes	19 autosomes + X/Y
Human	3.2 Gb	21000 protein coding genes	22 autosomes + X/Y

Human (MBC p206):
Mean gene size: 27 Kb
Mean number of exons: 10.4
Mean exon size: 145 nt (largest 17 Kb)

2 Modelling : a choice of description

The classical dogma of molecular biology is that genes are transcribed into mRNAs and that mRNA are then translated into proteins. Each step is actually a complicated process with many sub-steps and enzymes involved. There are of course also many differences between procaryotes and eukaryotes, the most obvious ones being the existence in eukaryotic cells of a nucleus and introns in genes. Moreover, with for instance the discovery of regulatory RNAs, the classical dogma itself appears now as only a part of the reality. We will here boldly start with a very simple description that hides most of these steps. We will then include these in some processes for which it is important to take them into account. The spirit is general to modelling. One should use the appropriate level of description bearing in mind that it is not useful and even detrimental to understanding to include inessential features. We use here ordinary deterministic differential equations to describe the kinetics as usually done in chemistry. It is intermediate between the binary "logic" description, in which species are present/absent and a stochastic description where events happens only with given probabilities. The binary description is somewhat simpler specially when it is associated to discrete time but then often not well suited to describe dynamic processes which will be our main focus. The stochastic description is required when noise plays a crucial role but is more complicated and often not useful, at least in a first approach.

3 Chemical kinetics/ODE representation of transcription and traduction

The simplest description of transcription of a gene g is simply that it produces a mRNA m at a rate ρ_m. We therefore describe this as

$$\frac{dm}{dt} = \rho_m - \delta_m m \tag{1}$$

At this stage, we do not describe explicitly maturation of mRNA and simply take m to simply represent mature mRNA for eukaryotes. We have described mRNA degradation as a linear process taking place at a rate δ_m.
Our mathematical account of translation is as simple. The production and

degradation of protein P from mRNA m is described as

$$\frac{dP}{dt} = \rho_P - \delta_P P \tag{2}$$

where P is the concentration of protein P, ρ_P is the protein production rate and δ_P its degradation rate. Translation from mRNA that have a concentration m, production is most simply summarized by writing

$$\rho_P = \beta m \tag{3}$$

That is, each mRNA molecule produces β proteins per unit of time.
To better see what the simple Eqs.(1,2,3) mean, let us suppose that the gene g was not active at $t < 0$, so that $m(t) = P(t) = 0$, $t < 0$. Its transcription is turned on at time $t = 0$ at the rate ρ_m. Then Eq. (1) gives:

$$m(t) = \frac{\rho_m}{\delta_m}[1 - \exp(-\delta_m t)], \quad t > 0 \tag{4}$$

A few simple comments can be made. What governs the approach to this asymptotic value is the degradation rate δ_m (as is already clear from dimensional analysis since $1/\delta_m$ is the only constant in the equation which has the dimension of time). It i usual to introduce the mRNA half-life $t_{1/2}$ (which measures the time after which the concentration is halved when one stops the production)

$$t_{m,1/2} = \frac{\ln(2)}{\delta_m} \simeq \frac{0.7}{\delta_m} \tag{5}$$

For times large compared to $t_{m,1/2}$, m is equal to its asymptotic value, ρ_m/δ_m, which is a simple balance from production and degradation. Therefore, the faster one would like to approach the steady value, the larger should the degradation rate be. If one likes to keep the steady value fixed, one should concomittantly increase the production rate ρ_m. This is costly and wasteful since one only produces molecules to degrade them. We will see later how adding some interactions can provide a remedy.
Given the time course of mRNA concentration $m(t)$ (Eq. (4)), the time course of the protein concentration is easily obtained by solving Eq. (3). Remember that the general solution of an inhomogeneous linear equation is the sum of a general solution of the homogeneous equation (here $A\exp(-\delta_P t)$), plus a particular solution of the inhomogeneous equation (here we try the same

functional terms as in m(t), $B + C\exp(-\delta_m t)$). Once the unknown constant is fixed by the boundary condition, $P = 0$ at $t = 0$, one obtains

$$P(t) = \frac{\beta\rho_m}{\delta_m\delta_P}\left\{1 - \exp(-\delta_P t) - \frac{\delta_P}{\delta_m - \delta_P}[\exp(-\delta_P t) - \exp(-\delta_m t)]\right\} \quad (6)$$

The protein concentration $P(t)$ starts quadratically at short times since its production rate increases linearly at short times, proportionally to the mRNA concentration $m(t)$.

Although, Eq. (1) and (2) already provide a very compact description of many elementary processes, it is sometimes inessential to describe these two steps separately. This can even be done with some justification if one of the two steps is much faster than the other. For instance, if mRNA dynamics is fast compared to protein dynamics (basically $\delta_m \gg \delta_P$ for all proteins involved), in a quasistatic approximation $m = \rho_m/\delta_m$. One can then describe the dynamics of P production with the single Eq. (2) with

$$\rho_P = \beta\rho_m/\delta_m. \quad (7)$$

Note that solving Eq. (2 with (7) agrees with directly taking the limit in $(\delta_P/\delta_m \to 0)$ in Eq. (6). Note also that some short time features of the kinetics of $P(t)$ are lost in the limit, like its quadratic initial increase.

The values of kinetics constants vary widely depending on the types of cells and the growth condition but it is useful to have some typical orders of magnitude. Several recent experiments have tried to measure them on a wide scale. Fig. 1 show the results of a recent experiment of Schwanhaüsser et al. [**?**] that used pulse labelling to monitor mRNA and proteins half-life on a genome wide scale in mouse fibroblasts growing in culture. Sequencing and mass spectroscopy were used to obtain mRNA and protein abundances. mRNA half-lifes were found to range from 1 hr to a day with a median value of 9hr (other studies report shorter values), while those of proteins were found to range from a few hours to a few days, with a median value of about 2 days. The mRNA numbers were found to range from 1 to a few tens per cell while proteins range from about hundred to 10^5 copies. Correspondingly production rates were found to be a few per hour for mRNA (our ρ_m denoted by ν_{sr} in [**?**]) with a few hundreds (median 180) of proteins produced per mRNA per hour (our β denoted by k_{sp} in [**?**]).

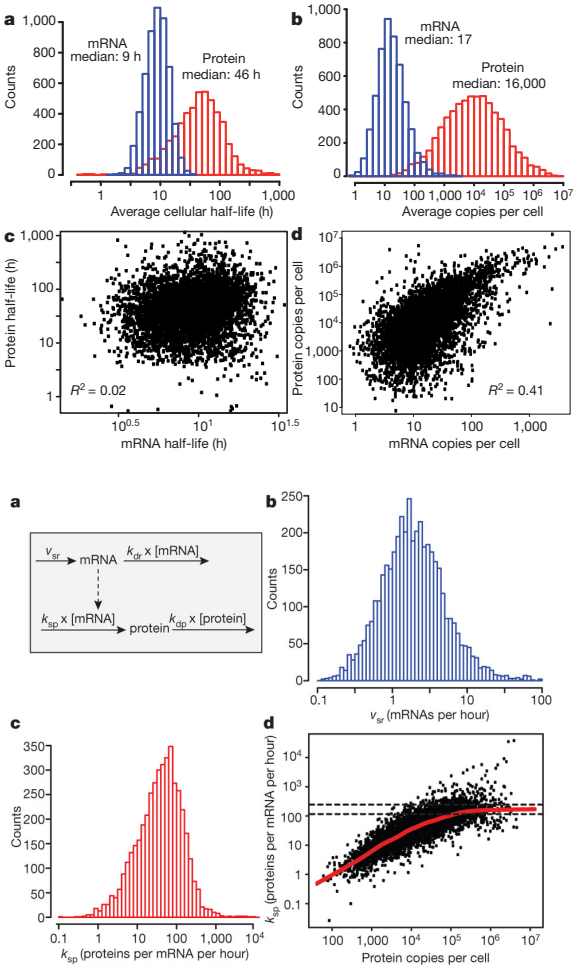

Figure 1: (A) a, b, Histograms of mRNA (blue) and protein (red) half-lives (a) and levels (b). Proteins were on average 5 times more stable and 900 times more abundant than mRNAs and spanned a higher dynamic range. c, d, Although mRNA and protein levels correlated significantly, correlation of half-lives was virtually absent. (B) a, mRNAs are synthesized with the rate vsr and degraded with a rate constant kdr. Proteins are translated and degraded with rate constants ksp and kdp, respectively. b, Calculated mRNA transcription rates show a uniform distribution. c, Calculated translation rate constants are not uniform. d, Translation rate constants of abundant proteins saturate between approximately 120 and 240 proteins per mRNA per hour. Red line shows the locally weighted fit (Lowess). Dashed lines indicate 95% confidence intervals of the Lowess maximum value calculated by bootstrapping. UNCORRECTED CAPTION/FiGURE

4 Regulation of transcription

One crucial element in gene networks is that some proteins can bind DNA, directly or indirectly by binding other proteins attached to DNA, and regulate the transcription of specific genes. A starting description of this process can be given as follows. We denote by T a transcription factor that regulates gene g and consider the two states, G, when T is not attached to a DNA region that regulates gene g, and TG, when on the contrary T is bound on a regulatory region and affects the transcription rate of gene g. The transition between these two states can be described schematically by the reaction

$$\mathrm{T} + \mathrm{G} \rightleftharpoons \mathrm{TG}$$

At equilibrium, one has

$$\frac{[T][G]}{[TG]} = K_d \tag{8}$$

Conservation of gene g between the two states can be written

$$[G] + [TG] = 1 \tag{9}$$

where we have measured concentrations in molecule numbers (by taking the cell -or its nucleus?-as unit volume). The 1 on the r.h.s. then arises from the fact that we have supposed that there is one copy of gene g. Finally, fractional values of $[G]$ or $[TG]$ can appear strange since T is either bound or not bound. They should be interpreted as the fraction of time that is spent in one or the other state. Eq. 9 can be used to express $[G]$ as a function of $[TG]$. The equilibrium Eq. (8) then gives,

$$[TG] = \frac{[T]}{K_d + [T]} \tag{10}$$

It is a usual assumption to suppose that the equilibrium binding of T to the promoter of gene g controls the action of T on the transcription of gene g. In this model, one then ascribes two different production rates for the mRNA m in the two states of the promoter of g (without or with T bound)

$$\rho_m(T) = \frac{\rho_0 K_d + \rho_T [T]}{K_d + [T]} \tag{11}$$

The corresponding curve is displayed in Fig. 2. Depending on whether ρ_T is

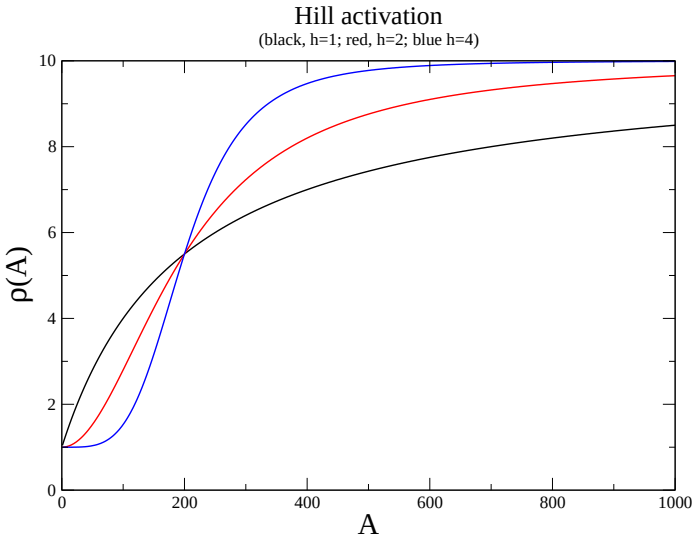

Figure 2: mRNA production rate ρ as a function of the activating transcription factor concentration T. The three curves corresponds to three different Hill coefficients, $h = 1$ (black, Eq. (11)), $h = 2$ (red,Eq. (12)) and $h = 4$ (blue, Eq. (12)).

larger or smaller than ρ_0, the factor T is an activator or a repressor of the transcription of gene g. Half-activation (or repression) is obtained for $[T] = K_d$. Typical values of K_d corresponds to a few tens of transcription factor molecules in a cell (nucleus) or nanomolar concentrations. Eq. (11) represents the simplest non-cooperative type of activation/repression of transcription of a gene. Before describing a few experiments that have measured this function (or the aggregated version of it Eq. (7)), it is worth giving the generalized version

$$\rho_m = \frac{\rho_0 + \rho_T([T]/K_d)^h}{1 + ([T]/K_d)^h} \tag{12}$$

The exponent h is called the Hill coefficient and values of $h > 1$ are usual. They generally describe in an effective manner cooperative activation of transcription by several molecules. We provide below two simple examples to show how dimer formation or cooperative binding of DNA can give rise to values of larger than 1. Values $h > 1$ are important because in this case K_d plays the role of a threshold. For concentration below K_d there is not much activation/repression, whereas for concentration above K_d activation/repression is almost maximal. This is most easily seen in the limit $h \gg 1$ in which the activation function of Eq. (12) reduces to a simple step function. Namely, for $h \gg 1$, when $([T] < K_d$, $([T]/K_d)^h$ is approximately zero since it corresponds to a number smaller than 1 raised to a large power, and, on the contrary, $([T]/K_d)^h$ is for $([T] < K_d$. Therefore, for $h \gg 1$ one has

$$\rho_m \simeq \rho_0 + (\rho_T - \rho_0)\theta([T] - K_d) \tag{13}$$

where we have introduced the Heaviside step function $\theta(x)$, $\theta(x) = 0$ for $x < 0$ and $\theta(x) = 1$ otherwise.

4.1 Sequential attachment

How can Hill coefficients greater than one appear? The basic idea is cooperativity. We give a simple example. We assume that two T's can sequentially bind the promoter of gene g. Namely, that one has the two reactions

$$\mathrm{T} + \mathrm{G} \rightleftharpoons \mathrm{TG} \tag{14}$$

$$\mathrm{T} + \mathrm{TG} \rightleftharpoons \mathrm{GT}_2 \tag{15}$$

with the equilibrium constants

$$\frac{[T][G]}{[TG]} = K_1, \ \frac{[T][TG]}{[T_2G]} = K_2 \tag{16}$$

Simple algebra gives that the fractional occupancy of the two T's bound species is

$$[T_2G] = \frac{[T]^2/(K_1K_2)}{1 + [T]/K_1 + [T]^2/(K_1K_2)} \tag{17}$$

The concentration at which the quadratic term becomes significant in Eq. (17) is $T^* = \sqrt{K_1K_2}$. The linear term in the numerator is negligible at this concentration when $K_2 < K_1$ This condition simply means that the concentration of T at which the second T binds is much smaller than the first, so that the binding of one T protein promotes the binding of a second one. As a result, the promoter is occupied by no T or 2. When this condition is satisfied, the binding and thus the activation of the promoter can be approximately described by a Hill function with a Hill exponent close to two, as shown in Fig. 3

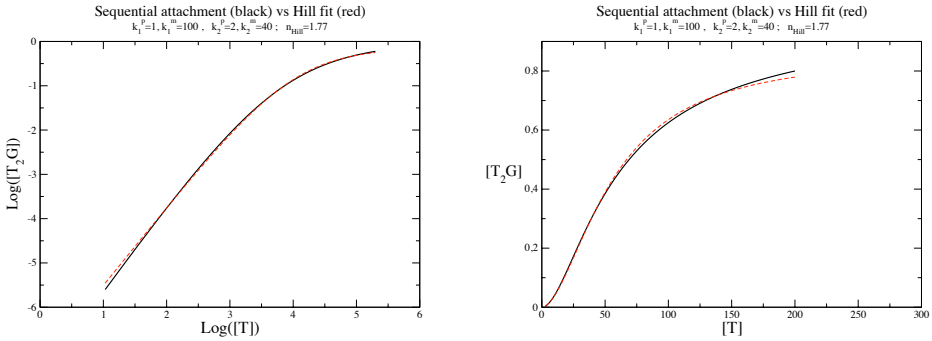

Figure 3: Sequential attachment can produce an effective Hill activation function. Left: parameter fit in Log-lin scale. Right: comparison of the activation function (red) and its Hill-function fitted form (dashed black line)

5 Binding of multiple factors : more complicated promoter functions

Transcription of a gene can be regulated by multiple transcription factors, and this seems even to be the norm in eucaryotic cells. The transcription rate can then be a complex function of the transcription factor concentrations. In the boolean approximation a gene regulated by an activator A and a repressor R can for instance be active only when A is present and R absent. Namely, it can be governed by the boolean function (A AND NOT R). Within our thermodynamic assumption, this can for instance be realized by the activator and repressor competing for the same binding sequence,

$$\mathrm{A} + \mathrm{G} \rightleftharpoons \mathrm{AG} \tag{18}$$

$$\mathrm{R} + \mathrm{G} \rightleftharpoons \mathrm{RG} \tag{19}$$

with the equilibrium conditions

$$\frac{[A][G]}{[AG]} = K_A, \ \frac{[R][G]}{[RG]} = K_R \tag{20}$$

The conservation equation $[G] + [AG] + [RG] = 1$, then gives the fractional abundance of the three states of the promoter

$$[G] = \frac{1}{1 + [A]/K_A + [R]/K_R}, \ [AG] = \frac{[A]/K_A}{1 + [A]/K_A + [R]/K_R}, \ [RG] = \frac{[R]/K_R}{1 + [A]/K_A + [R]/K_R}.$$

Assuming as before that transcription proceeds at different rates ρ_0, ρ_A, ρ_R in these three states one obtains the transcription rate

$$\rho([A],[R]) = \frac{\rho_0 + \rho_A[A]/K_A + \rho_R[R]/K_R}{1 + [A]/K_A + [R]/K_R} \tag{21}$$

An (A AND NOT R)-activation function is approximately realized by taking the basal rate and the repressed rates equal to zero, $\rho_0 = \rho_R = 0$.

6 An alternative description of gene activation: the thermodynamic formalism

The supposition that binding/unbinding of transcription factors to their cognate sites on DNA is fast compared to other processes (as we have done when

we used the Hill function) naturally leads to assume that their binding is at thermodynamic equilibrium. This provides a way to directly relate binding energies and DNA sites occupation.
The basic assumption is that if the binding energy for a transcription factor T on a DNA site is ϵ, the probability of T to be bound on DNA in presence of a solution containing T is proportional to $\exp[-\beta(\epsilon-\mu)]$ where μ is the chemical potential of T in the solution

$$\exp(\beta\mu) = [T]/T_0 \tag{22}$$

For instance for a single site, the two possible states are that the site is free and or the site is bound. Their respective probabilities are proportional to 1 and $\exp[-\beta(\epsilon-\mu)$. So the probability of the DNA site is occupied by T reads

$$p = \frac{\exp[-\beta(\epsilon-\mu)}{1+\exp[-\beta(\epsilon-\mu)} = \frac{[T]/K_d}{1+[T]/K_d} \tag{23}$$

with $K_d = T_0\exp(\beta\epsilon)$. This corresponds exactly to our previous description Eq. (**??**) with the bonus that we can in principle relate variation in K_d to variation in binding energies (the latter however are rarely directly available).
We can further extend the description by including the RNA polymerase. There are then 4 states of DNA, E(empty), with T, with R, with T and R. We take the empty state as our reference energy (0) and the energy of the three others state are ϵ_T, ϵ_R and $\epsilon_{T,R}$. We assume that the RNA pol binds the gene promoter with energy ϵ_R and the transcription factor with an energy $\epsilon_{T,R}$. The corresponding relative probablilities of the four states are then

$$p_E = 1,\, p_T = \exp[-\beta(\epsilon_T-\mu)],\, p_R = \exp[-\beta\epsilon_R], p_{T,R} = \exp[-\beta(\epsilon_{T,R}-\mu)] \tag{24}$$

The probability of the states R and T-R are

$$p_R = \frac{\exp[-\beta\epsilon_R]}{1+\exp[-\beta\epsilon_R]+\exp[-\beta(\epsilon_T-\mu)]+\exp[-\beta(\epsilon_{T,R}-\mu)]} \tag{25}$$

$$p_{T,R} = \frac{\exp[-\beta(\epsilon_{T,R}-\mu)]}{1+\exp[-\beta\epsilon_R]+\exp[-\beta(\epsilon_T-\mu)]+\exp[-\beta(\epsilon_{T,R}-\mu)]} \tag{26}$$

If one postulates that transcription rates are proportional to the promoter occupancy by RNA pol, one obtains for the transcription rates ρ_R of the

naked promoter and ρ_T with the factor T bound to DNA,

$$\rho_R = \frac{\rho \exp[-\beta\epsilon_R]}{1+\exp[-\beta\epsilon_R]+[T](\exp[-\beta\epsilon_T]+\exp[-\beta\epsilon_{T,R}])/T_0} \quad (27)$$

$$\rho_{T,R} = \frac{[T]}{T_0}\frac{\rho \exp[-\beta\epsilon_{T,R}]}{1+\exp[-\beta\epsilon_R]+[T](\exp[-\beta\epsilon_T]+\exp[-\beta\epsilon_{T,R}])/T_0} \quad (28)$$

where ρ is a reference transcription rate.
This is again the previous formula (11) with

$$\rho_0 = \rho\frac{\exp(-\beta\epsilon_R)}{1+\exp(-\beta\epsilon_R)},\ \rho_T = \rho\frac{\exp(-\beta\epsilon_{T,R})}{\exp(-\beta\epsilon_T)+\exp(-\beta\epsilon_{T,R})} \quad (29)$$

$$K_d = T_0\frac{1+\exp(-\beta\epsilon_R)}{\exp(-\beta\epsilon_T)+\exp(-\beta\epsilon_{T,R})} \quad (30)$$

The thermodynamic formalism is a way to rephrase our previous analysis. Not much is gained for simple promoters in the absence of direct information on binding energies (i.e the number of unknown parameters is the same). However, it leads to simple hypotheses and can bring some interesting information when transcription result from the combination of multiple transcription factor bindings (see e.g. Gertz-Siggia-Cohen).

7 An illustrative example: self-activation and bistability

7.1 A single self-activating gene

We will explore more systematically the dynamics of small networks in the next chapter. However, to illustrate the importance of transcription regulation characteristics, we consider here the simple case of a transcription factor that activates the transcription of its own gene, as depicted in Fig. 4. For simplicity, we do not describe mRNA dynamics.

$$\frac{dA}{dt} = \frac{\rho_0+\rho_1(A/A_0)^h}{1+(A/A_0)^h} - \delta_A A \quad (31)$$

What are the possible steady states of this system? In a steady state by definition the concentration remains constant and $dA/dt = 0$. Therefore the

steady state concentrations A_s obey

$$\frac{\rho_0 + \rho_1 (A_s/A_0)^h}{1 + (A/A_0)^h} = \delta_A A_s \tag{32}$$

For $h = 1$, Eq. (32) could be explicitly solved since it reduces to a finding the roots of a second-degree polynomial. The general result can be intuitively understood by plotting the the two sides of Eq. (32) as shown in Fig. 4. The steady concentrations are obtained when the two sides of Eq. (32) are equal, ie they are given by the abscissae of the intersection points of the two curves. For $h = 1$, there is a single intersection point. Moreover, it is stable and always reached at long times since for a concentration $A > A_s$, degradation of A is larger than production of A (the line corresponding to the r.h.s of Eq. (32) is above the curve for l.h.s.) anc ocncentartion decreases ($dA/dt < 0$). The reverse is true if $A < A_s$, and the same reasoning shows that A increases. When $h > 1$, the same is true as for sufficiently high degradation rates or sufficiently low degradtion rates. In these two cases, it is clear from Fig. 4 that the two curves have a single intersection point, and that it is stable by the same reasoning as above. However, in an intermediate range of degradation rates values, the two curves have three intersection points. The previous reasoning shows that the low and high concentration ones are stable while the intermediate one is unstable. Therefore, at long times whether the system stands at the high or low concentration points entirely depends on its history. This is the phenomenon of bistability : two cells with the same intermediate value of the degradation constant δ can stably be in two different states. The phenomenon has been directly observed in bacteria, as we discuss later and provides a simple starting conceptual model for differentiation. Auto-activation has in fact been reported for several "master genes" in development (scute, MyoD, Twist,...). Well-studied example : ComK in Bacillus subtilis autoactivate itself, threshold to jump from low to high state, bimodal in presence of noise (P Gamba, MJ Jonker, LW Hamoen, PLos Gen (2015)).

8 Two cross-repressing genes and bistability

Another simple small network that can display bistability is given by two genes that cross repress each other. This should be intuitively clear since repressing a repressor is an effective auto-activation. In order to analyze

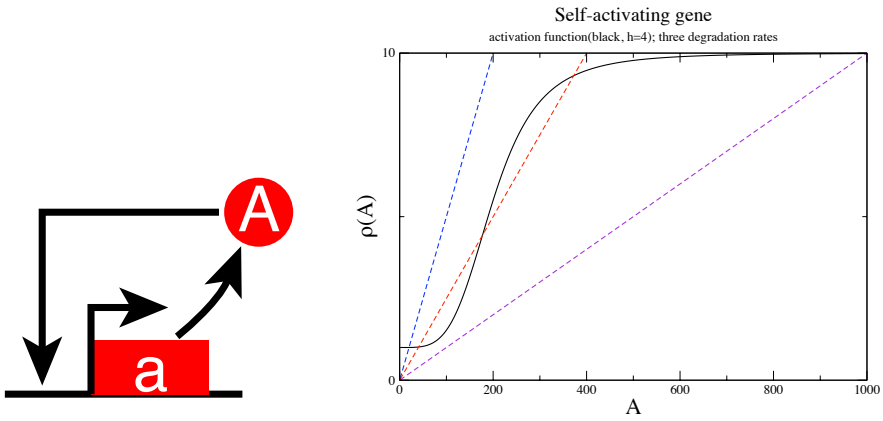

Figure 4: The simplest "network": protein A regulates the transcription of its own gene a (left). The right graph shows the gene-activation/protein-production function for a Hill exponent $h > 1$ (drawn here for $h = 4$, solid black line) as well as the protein degradation functions (color dashed lines for three values of the degradation rate. For a high degradation rate (dashed blue line) there is a single stationary state with a low concentration of A, for a low degradation rate (dashed violet line), there is a single stationary state with a high concentration of A. However for a range of intermediate degradation rates (dashed red line), there are three stationary states: the high and low concentration fixed point are stable and separated by an unstable fixed point at intermediate concentration.)

it mathematically, we consider the network between gene A and gene B described by the following set of equations

$$\frac{dA}{dt} = \rho_A(B) - \delta_A A \tag{33}$$
$$\frac{dB}{dt} = \rho_B(A) - \delta_B B \tag{34}$$
$$\tag{35}$$

with the repression function ρ_X with Hill kinetics

$$\rho_X = \frac{\rho_X}{1 + (X/X_0)^{h_x}} \tag{36}$$

The nullclines are given by the two graphs $A = \rho_A(B_1)/\delta_A$ and $B_2 = \rho_B(A)/\delta_B$. These two graphs always meet and there is a t least a steady state (see Fig. **??**) [1] It is easy to write the criterion for a steady state (of Eq. (37,38) to be stable. Linearizing these two equations as $A = A_s + A_1, B = B_s + B_1$, one obtains,

$$\frac{dA_1}{dt} = \rho'_A(B_s)B_1 - \delta_A A_1 \tag{37}$$
$$\frac{dB_1}{dt} = \rho'_B(A_s)A_1 - \delta_B B_1 \tag{38}$$
$$\tag{39}$$

The trace of this linear system is $-(\delta_A + \delta_B)$ is negative. The only condition for the point to be stable is that the determinant is positive, namely

$$\delta_A\delta_B - \rho'_A(B_s)\rho'_B(A_s) > 0, \text{ or equivalently } \rho'_B(A_s)/\delta_B > \delta_A\rho'_A(B_s) \tag{40}$$

Note that both derivatives ρ' are negative since A and B are repressors (Eq. (36)). Geometrically, Eq. (40) means that the slope of the B_1 graph is less negative that the slope of the B_2 at the intersection point. The two case of a stable intersection and an unstable one are shown in Fig. 5, together with some representative trajectories.

[1] It is not difficult to present a formal argument based on the finite maximal activation of the two genes. Consider the difference $\Delta B = B_1 - B_2$. When $A \to 0$ then $\Delta B \to +\infty$ (as B_1, since B_2 is finite as $\rho_B(0)$) while $\Delta_B < 0$ when $A_m = \rho_A(B_1 = 0)$ (since $B_2 = \rho_B(A_m)$ is positive at this point. ΔB being a continuous function of A there is a -steady state-point where $\Delta B = 0$ where the two curves meet.

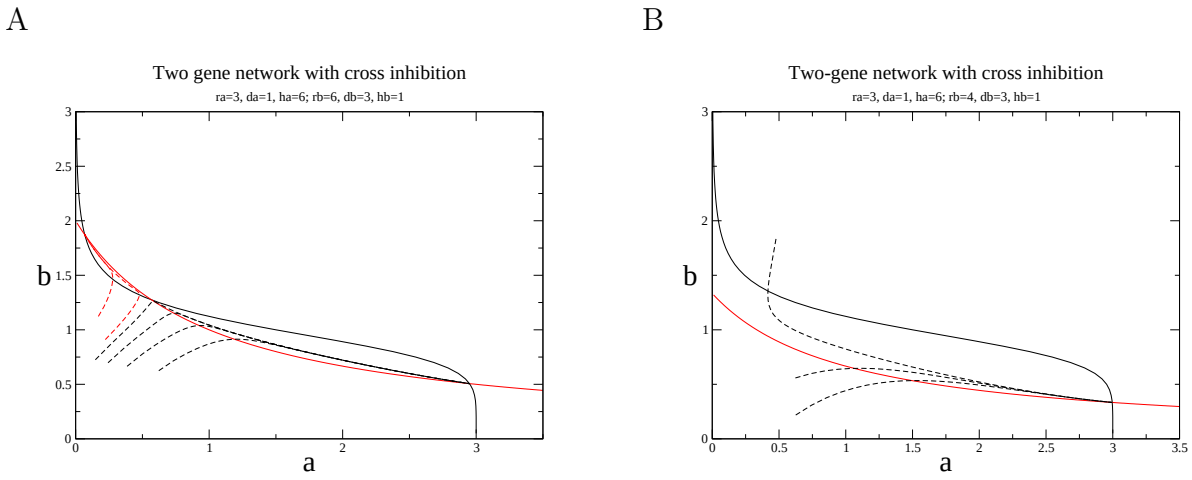

Figure 5: Two cross-repressing genes. (A) The single intersection point is stable. (B) The middle intersection point is unstable, the two other ones are stable.

9 Oscillation from self-repression with a delay.

Gene networks are observed to be oscillatory in several cases such as for the production of circadian rhythms. This is also the case during development, one famous example being somitogenesis, the production of somites, repeated segments in vertebrates that are at the origin of vertebrae, ribs,skeletal muscles,... They originate from oscillating transcription in the presomitic mesoderm. It has been proposed that direct autopression of the gene (her7 in zebrafish/Hes7 in mouse) is at the root of this segmentation clock [**?**, **?**]. We explain this mechanism here.

Let us consider a protein that represses the transcrition of its own gene. In order to study a simple model, we do not explicitly describe mRNA production,

$$\frac{dP}{dt} = \Phi[P(t)] - \delta P(t) \tag{41}$$

We will keep Φ general for the analysis below but we have in mind something

like Hill repression with a coefficient h e. g

$$\Phi[P] = \frac{\rho_0}{1 + (P/P_0)^h} \tag{42}$$

Φ describes the self-repression by P of its own gene. The steady state is the solution of $P_s = \Phi[P_s]$. It is easy to see graphically that it is unique when Φ is decreasing, as shown in Fig. 6 To investigate the stability of this steady solution, we write $P(t) = P_s + P_1(t)$ and suppose that P_1 is a small departure from P_s, $|P_1| \ll P_0$. We can thus linearize, Eq, (44) and obtain the equation that controls the dynamics of the perturbation P_1,

$$\frac{dP_1}{dt} = \Phi'[P_s]P_1 - \delta P_1 \tag{43}$$

Since Φ is decreasing the two linear terms are negative and P_s is stable.
However, when one takes into account delays between gene activation and protein production, iself-repression can give rise to oscillations (that negative feedback with delay can produce oscillations can be experienced when using a shower with a slow response to temperature adjustment). We modify the previous model and write

$$\frac{dP}{dt} = \Phi[P(t - D)] - \delta P(t) \tag{44}$$

where D is a delay. It is a simple way to take into account the delay between the activation of the promoter and the actual appearance of the protein. Proteins that appear at t depend on the state of the gene promoter at $t - D$, which itself depends on protein concentration at $t - D$ if binding transcriptional activation/repression are fast. Normalizing P by P_0 and time by $1/\delta$, it is easy to see that the dynamics is controlled by the two dimensionless parameters δD and $\rho_0/(P_0\delta)$ (as well as the Hill coefficient h). We will apply the general formulas to this case fro definiteness.

Figure 6: Self-repressing gene and oscillations.

To investigate the stability of this steady solution, we write, as before, $P(t) = P_s + P_1(t)$ and suppose that P_1 is a small departure from P_s, $|P_1| \ll P_0$.

We can thus linearize, Eq, (44) and obtain the equation that controls the dynamics of the perturbation P_1,

$$\frac{dP_1}{dt} = \Phi'[P_s]P_1(t-D) - \delta P_1 \tag{45}$$

The steady solution P_s is stable if all perturabtions decreases as time grows. In order to investigate when this is the case, we search for the eigenmodes of Eq. (45) under the form $P_1(t) = A\exp(\sigma t)$. This gives the characteristic equation that determines the eigenvalues σ,

$$\sigma = -K\exp(-\sigma D) - \delta \tag{46}$$

with the constant $K = -\Phi'[P_s] > 0$ (we have explicitly introduced the minus sign to have a positive constant). When K is small, one solution is close to $-\delta$ and the other ones should have a large (in absolute value) and negative real part so that the exponential term cannot be neglected. So, for K small enough, all eigenvalues σ) have a negative real part and perturbations decreases as time grows : P_s is stable. If this cease to be true as K grows, at some point the real part of some eigenvalues becomes positive. Thus, at this threshold K_c, Eq. (46) should have a purely imaginary solution $\sigma = i\omega_c$, with $-K_c\exp(-i\omega_c D) = \delta + i\omega_c$.
Separating real and imaginary parts gives

$$K_c = -\frac{\delta}{\cos(\omega_c D)} = \sqrt{\delta^2+\omega^2},\ \tan(\omega_c D) = -\omega_c/\delta$$

Parametric solution

$$\frac{\pi}{2} < a \leq \pi,\ \omega_c = a/D,\ \frac{1}{D\delta} = -\frac{1}{a}\tan a,\ K_c = -\frac{\delta}{\cos(a)} \tag{47}$$

When $1/(D\delta) \to 0, K_c \to 1$ (limit $a \to \pi$)
(i.e. for long delays, the period $T = 2\pi/\omega_c = 2\pi D/a = 2D$
When $1/(D\delta) \to +\infty, K_c \sim \pi/(2D)$ (limit $a \to \pi/2$)
(i.e for short delays, $T = 2\pi/\omega_c = 2\pi D/a = 4D$). The critical slope K_c as well as the frequency and period at threshold are drawn in Fig. 7 as a function of the delay.
When Φ is described by a Hill function with coefficient h (Eq. (42), it is easy to translate the above analysis in terms of the physical (adimensionned) parameters $\tilde{\rho} = \rho_0/(P_0\delta)$ and $\tilde{D} = D\delta$. With $Q = P/P_0$, the dynamics reads

$$\frac{dQ}{dt} = \frac{\tilde{\rho}}{1+[Q(t-\tilde{D})]^h} - Q(t) \tag{48}$$

where t is here measured in unit of $1/\delta$ but we keep the symbol t for simplicity. The fixed point Q_s satisfies

$$Q_s(1+Q_s^h) = \tilde{\rho} \tag{49}$$

The derivative $-K$ the fixed point reads

$$K = \tilde{\rho}\frac{h\,Q_s^{h-1}}{(1+Q_s^h)^2} = \frac{h\,Q_s^h}{(1+Q_s^h)} \tag{50}$$

where we have used Eq. (49) to eliminate $\tilde{\rho_s}$ in the second equality. This shows that K varies from 0 to h when Q_s varies from 0 to $+\infty$. Therefore, the linear oscillatory instability requires $h > 1$. Eq. (50) can be used to express Q_s and $\tilde{\rho}$ in terms of K and therefore to transcribe our previous stability results. Namely,

$$Q_s = \left[\frac{K}{h-K}\right]^{1/h} \tag{51}$$

$$\tilde{\rho} = \left[\frac{K}{h-K}\right]^{1/h}\frac{h}{h-K} \tag{52}$$

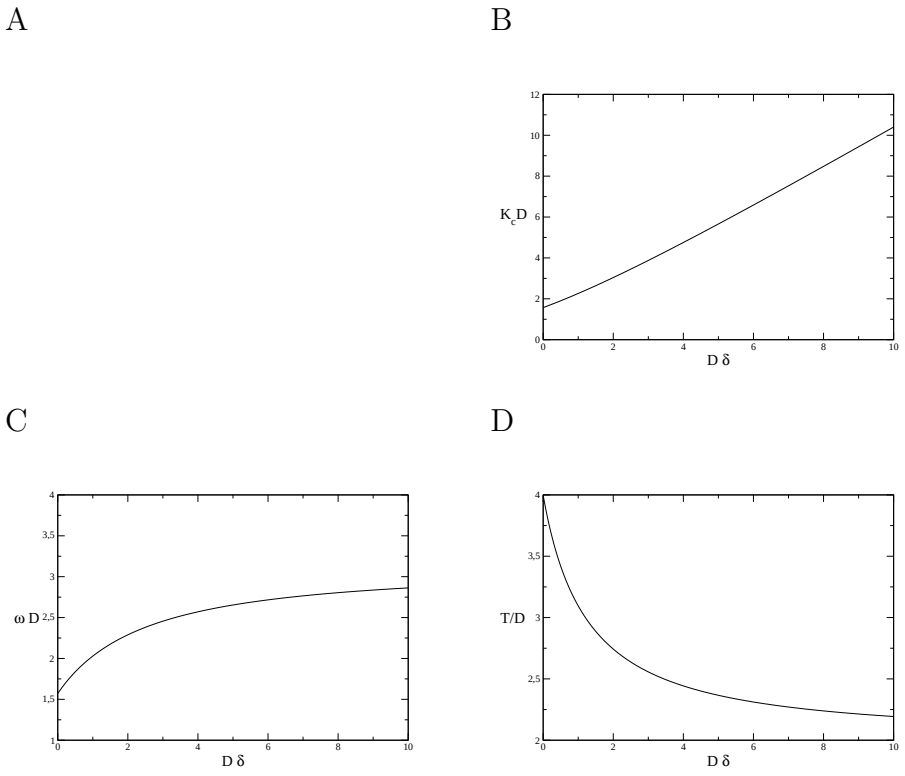

Figure 7: The self-repressing gene. (A) Stability boundary in the $(D\delta, K_c/\delta)$ plane. The fixed point is linearly unstable above the solid line. Note that K_c/δ is always larger then 1 (B) The stability boundary in the $(D\delta, K_cD)$ plane. It can be seen that for small delays $K_c \sim \pi/(2D)$. (C, D) The frequency and period of the oscillations at threshold as a function of the delay.