Math primer for the perplexed "gene-expression"-ist

Nitzan Rosenfeld

Weizmann Institute of Science

Rehovot , Israel

Cell growth[1]

First things first. The cells grow and divide, the biomass (N) increases. How quickly? If we assume that the amount of new biomass produced (dN/dt) is proportional to the amount of biomass present (N) – not a very far-fetched idea, as life begets life – we can write

$$\frac{dN(t)}{dt} = \alpha \cdot N(t) \quad .$$

The solution is just the familiar exponential growth,

$$N(t) = N_0 \cdot e^{\alpha \cdot t} \quad .$$

Note that if the biomass (or the number of cells) is doubled every time interval of τ , then this means that $e^{\alpha \cdot \tau} = 2$ or that $\alpha = \ln(2)/\tau$.

cell cycle	proteins per cell
0	8
1	4
2	2
3	1
∞	0
$\infty+1$	0

Protein Dilution[1]

Alright, so the biomass increases. How does this affect protein concentrations? It means that a protein that isn't actively produced will eventually be diluted out by growth – see figure where initially a cell has 8 copies of the protein, and eventually these 8 copies are spread out over its many progeny.
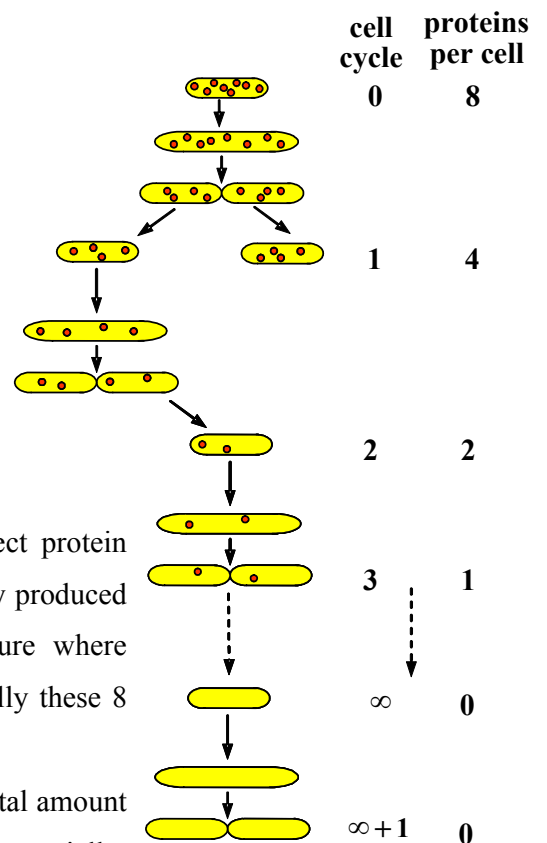
For a stable protein which isn't actively produced, the total amount of protein (X) is constant, while the biomass grows exponentially. The concentration of the protein (x) in the biomass is equal to

$$x(t) = \frac{X(t)}{N(t)} = \frac{X_0}{N_0 \cdot e^{\alpha \cdot t}} = \frac{X_0}{N_0} \cdot e^{-\alpha \cdot t} = x_0 \cdot e^{-\alpha \cdot t}$$

which goes to zero for $t \gg \alpha$. Just for curiosity (not really; it will come in handy soon) we note that x(t) satisfies the differential equation

$$\frac{dx(t)}{dt} = -\alpha \cdot x(t) \quad ,$$

1

which can be interpreted to mean that the growth provides a "sink" term for the protein concentration (even though the total amount of protein remains constant).

Protein degradation

But hold on, what happens if the protein is not stable? The amount of protein degraded at each moment (dX/dt) is proportional to the amount of protein present at that time (X(t)):

$$\frac{dX(t)}{dt} = -\alpha_{deg} \cdot X(t)$$

thus with no production, the protein decays exponentially

$$X(t) = X_0 \cdot e^{-\alpha_{deg} \cdot t}.$$

When the protein is diluted out by growth as well as degraded, we obtain

$$x(t) = \frac{X(t)}{N(t)} = \frac{X_0 \cdot e^{-\alpha_{deg} \cdot t}}{N_0 \cdot e^{\alpha \cdot t}} = \frac{X_0}{N_0} \cdot e^{-\left(\alpha_{deg} + \alpha\right)t} = x_0 \cdot e^{-\alpha' \cdot t}$$

which is identical in form to the previous equation, with

$$\alpha' = \alpha_{deg} + \alpha,$$

i.e. a faster rate of exponential decay. Note that

$$\tau_{effective} = \frac{1}{\dfrac{1}{\tau_{deg}} + \dfrac{1}{\tau_{cell\text{-}cycle}}}$$

so that

$$\tau_{effective} = \tau_{cell\text{-}cycle} \text{ for } \tau_{deg} \gg \tau_{cell\text{-}cycle}$$

$$\text{and } \tau_{effective} = \tau_{deg} \text{ for } \tau_{deg} \ll \tau_{cell\text{-}cycle}.$$

From here on, we will treat all proteins as stable, and use only α for convenience of notation. All of the following is identical for degradable short-lived proteins, after a substitution of $\alpha' = \alpha_{deg} + \alpha$ instead of α everywhere. Things get more interesting if the degradation rate is not constant but can vary over time. But this is beyond the scope of this short primer.

Protein production[1]

So what happens when protein is produced? We assume homogeneity of the cell culture, and that the same amount of protein per unit time is produced by each cell, so that the total amount of protein produced depends on the biomass. Now would be a

good time to abandon the "whole-culture" perspective and to start looking at intracellular concentrations of proteins instead of whole-culture amounts of them.

cell cycle	proteins per cell
0	0
1	4
2	6
3	7
∞	8
$\infty+1$	8

The figure may help provide intuition into what is happening. Consider a cell which starts out with no protein. Suppose that production is abruptly turned on so that during one cell-cycle it produces 8 copies. But then it divides, giving 4 copies to each progeny. These now produce 8 copies each, ending up with 12 copies that they must divide amongst their progeny, who get 6 copies each. They then produce 8 more, ending up with 14, giving 7 to each daughter cell, etc. One can see that both production and dilution by growth take effect here. (Note that although the number of molecules in a cell may go up and down in this example, the concentration never decreases, since the daughter cells have half the volume of the mother cell.)
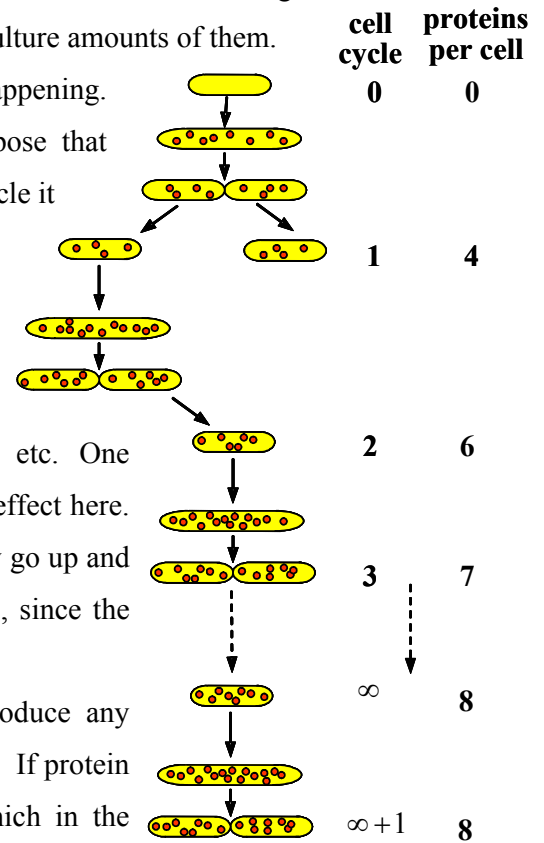
We already found out that for a cell that doesn't produce any protein, the protein concentration has a sink term $-\alpha \cdot x(t)$. If protein is also produced, then we must add a source term, which in the general case may vary over time so we denote it by $A(t)$ and write

$$\frac{dx(t)}{dt} = A(t) - \alpha \cdot x(t) \ .$$

To study induction, we shall assume the cell starts with no protein and then starts to produce protein at a constant rate A. To solve the differential equation with a constant rate of production $A(t)=A$, we guess a solution of the form

$$x(t) = C_1 + C_2 \cdot e^{-\alpha \cdot t} \rightarrow \frac{dx(t)}{dt} = -\alpha \cdot C_2 \cdot e^{-\alpha \cdot t} = -\alpha \cdot (x(t) - C_1) = \alpha \cdot C_1 - \alpha \cdot x(t).$$

Comparing this to our differential equation we see that $C_1 = A/\alpha$, and the solution has one free parameter (integration constant) determined by initial conditions:

$$x(t) = \frac{A}{\alpha} + C \cdot e^{-\alpha \cdot t} \ .$$

If we look at cases where the initial value is zero, $x(t=0)=0$, then $C = -A/\alpha$, and

$$x(t) = \frac{A}{\alpha} \cdot \left(1 - e^{-\alpha \cdot t}\right) .$$

Another way to write this is

$$x(t) = \frac{A}{\alpha} \cdot \left(1 - 2^{-t/\tau}\right)$$

(remembering that $\alpha = \ln(2)/\tau$).

3

Gene repression – the Michaelis-Menten curve[2]

We now leave the domain of biomass growth and cell reproduction to look at simple modeling of molecular interaction. Later we shall combine the two to derive interesting results. Let us start with a simple model of gene repression, in which we include the following species:

S – Promoter region of gene x on a DNA strand.

R – Repressor of gene x.

[SR] – Repressor bound to the promoter, thereby interfering with the binding of RNA polymerase.

In order to define the relation between these species in mathematical terms, we use the conservation rule for each of the species, which states that the total number of promoters (repressors) is conserved, and they can take on either of two forms: as free promoters **S** (repressors **R**) or as promoters bound to repressors **[SR]**. We thus obtain the conservation equations:

$$S + [SR] = S^{tot},$$

$$R + [SR] = R^{tot}.$$

We assume that the rate of transcription of gene x is proportional to the fraction of **S** which is not bound to **R** (this is equal to the fraction of time each promoter is unbound to repressor when averaging over long time period compared to characteristic time of the binding-unbinding process).

The reaction scheme is:
$$K_{on} \rightarrow$$
$$S + R \longleftrightarrow [SR]$$
$$\leftarrow k_{off}$$
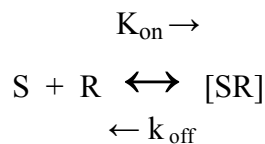
and the rate equation for change in the amount of complex **[SR]** is:

$$\frac{d[SR]}{dt} = k_{on} \cdot S \cdot R - k_{off}[SR].$$

We assume that after an initial transient, the reaction equilibrates and reaches a steady state in which the concentrations of the various species stops changing, meaning that

$$\frac{d[SR]^{st}}{dt} = k_{on} \cdot S^{st} \cdot R^{st} - k_{off}[SR]^{st} = 0 \quad \Rightarrow \quad k_{on} \cdot S^{st} \cdot R^{st} = k_{off}[SR]^{st}.$$

We now and applying the conservation equation to the steady-state equation to obtain

4

$$(1) \quad k_{on} \cdot S^{st} \cdot R^{st} = k_{off}[SR]^{st} = k_{off}(S^{tot} - S^{st}) = k_{off}(R^{tot} - R^{st}).$$

Up to this point, as you can see from the symmetry in the above equation, the promoters and the repressors are on equal footing. At this point it simplifies matters greatly to assume that one of the species is in large excess of the other. In our case we shall assume that the total amount of repressors is much greater than the total number of promoters, reflecting the fact that promoters are on the DNA and appear in only one or a few copies, while the repressors are proteins in the cells which appear in greater numbers. Putting this fact in mathematical terms, we write

$$R^{tot} \gg S^{tot}.$$

and since the amount of promoters bound to repressors cannot exceed the total number of promoters $S^{tot} \geq [SR]$, we can safely write that $R^{tot} \gg [SR]$ and therefore

$$R = R^{tot} - [SR] \approx R^{tot},$$

meaning that the amount of free repressors is barely affected by the number of promoters in the system (from now on we will ignore the difference between R and R^{tot}, and for clarity omit the suffix "^{tot}" from R^{tot}). Using this in equation (1) above, we obtain for the concentration of free promoters:

$$k_{on} \cdot S^{st} \cdot R = k_{off}(S^{tot} - S^{st}) \quad \Rightarrow \quad (\frac{k_{on}}{k_{off}}R + 1)S^{st} = S^{tot}$$

and defining $k_d \equiv \dfrac{k_{off}}{k_{on}}$, we obtain the neat form which is commonly referred to as the Michaelis-Menten equation, which gives the steady-state amount of free repressors as a function of the amount of repressors in the system:

$$(2) \quad S^{st}(R) = \frac{S^{tot}}{1 + \left(\dfrac{R}{k_d} \right)},$$

and the rate of transcription, which is proportional to S^{st}, is

$$A(R) = \frac{\beta}{1 + \left(\dfrac{R}{k_d} \right)},$$

where β is the amount of protein produced per unit time from S^{tot} unrepressed promoters. The rate of production decreases to zero when the amount of repressors is very large. Very large compared to what? Very large compared to k_d: for $R \gg k_d$,

$$S^{st}(R \gg k_d) \ll S^{tot} \ .$$

If the protein is an inducer rather than a repressor, then the production rate may be proportional to the amount of complex [SR] which we can also find from the model.

Negative autoregulation

What happens if a gene-product repressed the transcription of its own gene? If we neglect time required for protein synthesis and assume that at each moment the production rate of X, A(t), is a Michaelis-Menten-like function of its concentration at time t, x(t), we obtain the following differential equation:

$$(3) \quad \dot{x}(t) = \frac{dx(t)}{dt} = A(t) - \alpha \cdot x = \frac{\beta}{1 + x/k} - \alpha \cdot x$$

whose steady-state is (dx/dt=0):

$$x^{st} = \frac{\sqrt{k^2 + 4 \cdot k \cdot \beta/\alpha} - k}{2} \ .$$

We play around with eq. (3), writing

$$\frac{dx(t)}{dt} = \frac{\beta}{1 + x/k} - \alpha \cdot x \frac{1 + x/k}{1 + x/k} = \frac{\beta - \left(\alpha \cdot x + \alpha \cdot x^2/k\right)}{1 + x/k}$$

which we invert to obtain

$$dt = \frac{\left(1/k\right) \cdot x + 1}{-\left(\alpha/k\right) \cdot x^2 - \alpha \cdot x + \beta} \cdot dx \ .$$

This is a ratio of simple polynomial, whose integral is to be found in any mathematical handbook, giving the implicit solution of eq. (3), using x^{st} to simplify:

$$(4) \quad t(x) - t(x_0) = -\frac{1}{2 \cdot \alpha}\left[\log\left((x - x_{st})(x + k + x_{st})\right) + \frac{k}{k + 2x_{st}}\log\left(\frac{x - x_{st}}{x + k + x_{st}}\right)\right]_{x_0}^{x} \ .$$

For strong negative autoregulation, in the limit where $(\beta/\alpha) \gg k$, we approximate

$$x^{st} = \sqrt{k \cdot \beta_2/\alpha} \gg k$$

6

and simplify the right-hand side of eq. (4) to first order by neglecting the second term, which has a coefficient (k/x^{st}) much smaller than 1, and neglecting (k/x^{st}) in the first term:

$$t(x) - t(x_0) \approx -\frac{1}{2 \cdot \alpha} \left[\log((x - x_{st})(x + x_{st}))\right]_{x_0}^{x} = -\frac{1}{2 \cdot \alpha}\left[\log(x^2 - x_{st}^2)\right]_{x_0}^{x}$$

We now multiply by $-2 \cdot \alpha$ and expand the boundary conditions of the integral,

$$-2 \cdot \alpha \cdot (t(x) - t(x_0)) = \log(x^2 - x_{st}^2) - \log(x_0^2 - x_{st}^2) = \log\left(\frac{x^2 - x_{st}^2}{x_0^2 - x_{st}^2}\right),$$

and taking the exponent of both sides we obtain

$$(x_0^2 - x_{st}^2) \cdot \exp(-2 \cdot \alpha \cdot (t(x) - t(x_0))) = x^2 - x_{st}^2$$

which we can write explicitly as

$$x(t) = \sqrt{x_{st}^2 \cdot [1 - \exp(-2 \cdot \alpha \cdot (t - t_0))] + (x(t_0))^2 \cdot \exp(-2 \cdot \alpha \cdot (t - t_0))} \ .$$

For the case of induction, when $t_0 = 0$, $x(t_0 = 0) = 0$, we finally obtain the induction kinetics for strong negative autoregulation (in the limit where $(\beta/\alpha) \gg k$):

$$\frac{x(t)}{x^{st}} = \sqrt{1 - e^{-2\alpha \cdot t}} \quad , \quad x(t_0 = 0) = 0 \quad , \quad \beta/\alpha \gg k \ .$$

References:

1   Monod, J., Pappenheimer, A. M. & Cohen-Bazire, G. (1952). La cinetique de la biosynthese de la b-Galactosidase chez E. Coli consideree comme fonction de la croissance. *Biochimia et Biophysica Acta* 9, 648-660.
2   Stryer, L. (1995). *Biochemistry*. 4th edition, W. H. Freeman and Company, New York, pages 192-195.