Assignment 2: Regulatory & signalling pathways

Computational Biology 2020-2021

Deadline 21 April 2021

Introduction

This week we will take a first look at dynamics of regulatory and signaling pathways in cells. Here you will go further with solving ODEs, a common approach to model molecular networks. We will infer reaction rates and signal response curves of, among others, positive and negative feedback loops.

The exercises are ought to be made in python, except for the mathematical derivations. You are recommended to use the numpy and matplotlib package. Especially the numpy ODEINT-function may come in handy.

Quantitative chemical kinetic models

Biochemical and cellular processes follow chemical and physical laws. If underlying chemical processes are known, the behaviour of a system can be described on the basis of thermodynamical and chemical kinetics [1].

The most common approach to model molecular networks uses systems theory applied to chemical kinetics. Chemical kinetics uses the notion of processes where reactants are consumed and products are generated. The rate at which this process takes place is dependent on the quantity of the components involved, reaction order, and parameters. The mathematical expression depends on the assumptions made.

The emergent behaviour of a given system is from the combination of all the processes affecting the components: values of the variables, e.g. protein amounts or gene activities, can be affected by processes such as binding, catalysis, degradation etc.. The resulting effect on a component will be the sum of the rates of all processes influencing the component, multiplied by the stoichiometries of the processes.

Note that it is assumed in chemical kinetics that there is a homogeneous distribution of components in volumes. However, in reality most signalling reactions have only a few components that are also heterogeneously located in spatially complex compartments. For gene regulation it is ever more so, since there are very few partners and the reactions take place on complex folded nucleic acid (DNA).

Stoichiometric network

Two simple examples of protein dynamics are shown in Figure 1. The solid line and the dotted line respectively represent an chemical conversion and a dependence on a component. These networks involve a signalling strength S (e.g. concentration of mRNA) and a response element R (e.g. protein concentration).

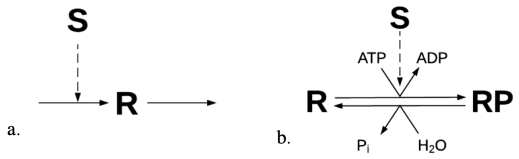

Figure 1: Two simple networks: a. Linear system, b. Loop system

The first model shows a **linear structure** (a), a simple model based on synthesis and degradation. The synthesis of R is dependent on S, e.g. mRNA. Note that we often see that the synthesis of a compound is never zero, even when there is no signalling strength. This is often modeled as a base concentration of mRNA.

The second model shows a **loop structure** (b) shows how the signalling strength S influences the phosphorylation of the response element. Phosphorylation of an enzyme means that a phosphoryl group is added to the enzyme, which often activates the enzyme. This model is especially interesting because it contains a **moiety conservation**, meaning that there is a molecule whose quantity is constant over time, but can be found in different forms or on different chemicals. We often refer to this quantity as the total concentration, which is a sum of compounds that contain (or are) this specific chemical part. A typical example would be the conserved moiety adenine in its forms ATP, ADP, AMP etc. In our example we can, next to others, identify a moiety conservation of R and R_p, with a total concentration of R_t. These conserved moieties can help solving stoichiometric models by reducing the number of **dynamic variables**. Our system of R and RP has two of them, but after replacing R with $R_t - R_p$, we obtain a system with only one dynamic variable (R_p).

Adding the dynamics

The next step is to transform the network into a dynamic system. This is done by defining the **rate equations** ($\frac{dR}{dt}$) in the system, which are ODEs (ordinary differential equations) determining the rate of change over time. We define a rate equation as the sum of all **reaction rates** (v_i) involving the dynamic variable. These reaction rates follow from the law of mass action, which states that we can write them down as the product of the concentration of all reactants multiplied with a **reaction constant** (k_i). Suitable parameters, or rather the lack thereof, is the bottleneck in building quantitative kinetic models using this modelling approach. One way to address this problem is to estimate the parameter values from experimental data sets. This estimation is a form of global optimization. We will look into parameter estimation and optimization later on in the course.

Our linear model has two reaction rates: synthesis (v_1) and degradation (v_2), respectively dependent on S and R. This results in the following rate equation with $\{k_0, k_1, k_2\} = \{1, 2, 3\}$, where k_0 represents the synthesis rate based on the base concentration of mRNA.

$$\frac{dR}{dt} = v_1 - v_2 = (k_0 + k_1 S) - (k_2 R)$$

The loop model has one reversible reaction, so our reaction rate can be split into two parts, a forward reaction rate (v_+) and a backwards reaction rate (v_-). The dependence on other reactants are assumed to be negligible, resulting in the following rate equation for R_p with $\{k_+, k_-, R_t\} = \{1, 1, 1\}$

$$\frac{dR_p}{dt} = v_+ - v_- = (k_+ S(R_t - R_p)) - (k_- R_p)$$

When solving these systems we use the **steady-state solution**, meaning that all concentrations are assumed to be constant.

Analyzing the dynamics

Plotting the system over time shows that both systems end up in an equilibrium, meaning that the values of the variables become constant over time, which is in biochemistry more commonly known as a **steady state**. We can identify two types of equilibria considering these kinds of systems. The first is a **thermodynamic equilibrium**, where there is no chemical activity left and all reaction rates v_i are equal to zero ($v_i = 0\ \forall i$). On the other hand we have the **dynamic equilibrium**. In this state of the system all concentrations are constant, but there is still chemical activity. It means that the sum of all producing reactions of a species is equal to its consuming reactions. In other words, the rate equations are all equal to zero and while there exists at least one nonzero reaction rate in the system ($\exists v_j \neq 0 : \frac{dR_i}{dt} = 0\ \forall i$). These equilibrium-concentrations are also referred to as equilibrium points, often indicated by an Asterisk (*). Figure 2 shows how both of the systems go to a equilibrium point, regardless of the initial concentration of R or R_p. Note that equilibrium point is dependent on S, meaning that R^* and R_p^* can be regulated by signal strength S.

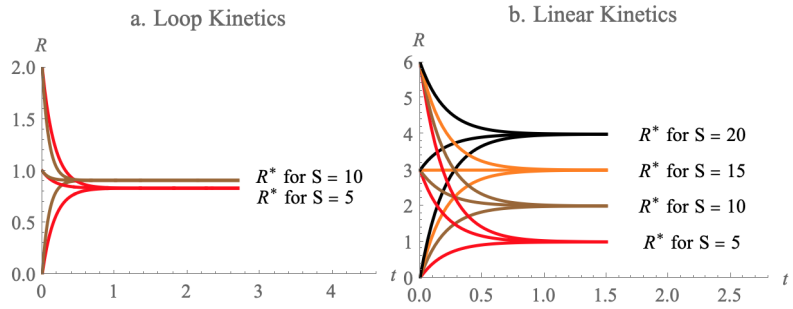

Figure 2: R and R_p over time for different initial values and S

By definition, we know that R is a constant value in equilibrium point R^*, so it can be concluded that its derivative, rate equation $\frac{dR}{dt}$, must be equal to zero. This allows us to find all equilibrium points of R by obtaining the roots of $\frac{dR}{dt}$. That is why plotting the rate equation ($\frac{dR}{dt}$) versus the concentrations of the corresponding compound (R) provides us with valuable information. It shows the values of the equilibrium points and it gives insight in the dynamics of the system for different values of R. Figure 3 shows the rate equations for our linear model. Imagine that we start at a certain value for R at the horizontal axis. Depending on the value of R, R will increase ($\frac{dR}{dt} > 0$), decrease ($\frac{dR}{dt} < 0$) or remain unchanged ($\frac{dR}{dt} = 0$). This change of R for different values of R is referred to as the **flow** of R, which goes to the right for $\frac{dR}{dt} > 0$ and to the left for $\frac{dR}{dt} < 0$, as indicated by the arrows. Only if the derivative is zero and the flow is zero, where the rate equation intersects the horizontal axis, we find ourselves in an equilibrium point. Systems can have multiple equilibrium points,

sometimes even an infinite amount. The flow round these points provides us with information on their **stability**. A **stable point** is an equilibrium point where the flow on both sides is directed towards it. These points are also referred to as sinks and are graphically indicated by a solid point. An **unstable point** is a point where the flow is away from it. This means that a small perturbation pushes the system out of its equilibrium. These points are also named sources and they are often indicated by a open points. Sometimes an equilibrium point is a **metastable point**, when flow goes towards it from one side and away from the other side. A quick way to check this is to take the derivative of the rate equation at these equilibrium points, which are respectively: negative, positive and zero.

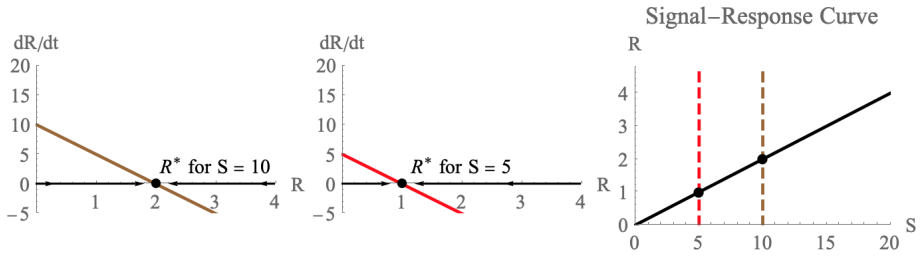

Figure 3: Linear kinetics: Rate equations and Signal-Response curve

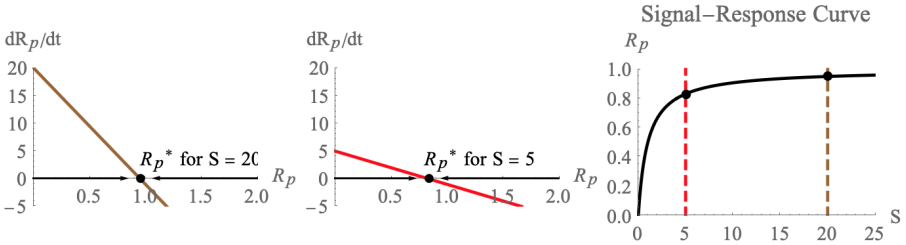

Figure 4: Loop kinetics: Rate equations and Signal-Response curve

How the equilibrium points relate to the signalling strength S can be shown in a **Signal-Response plot**, as shown in Figure 3 and 4. These plots show for different S: how many equilibrium points there exist, for which values they exist and they provide us with information on their stability (stable and unstable are respectively indicated by a solid and dashed lines). The main difference between the signal-response curves of our two systems originate from the conserved moiety in the loop kinetics. This limit on the total amount of R, R_{total}, creates a limit for R_p as S goes to infinity ($R^* \to 1$ as $S \to \infty$). This in contrast to R in the linear kinetics, where R increases linearly with S and is never contained ($R^* \to \infty$ as $S \to \infty$).

Bifurcation Points

Signal-Response curves allow us to see **bifurcation points**: points at which the stability of the system abruptly changes. This is why these signal-response plots are also known as **bifurcation diagrams**. Figure 5, 6 and 7 show the shape of three common bifurcations, respectively: the **Saddle-Node bifurcation**, where a set of equilibrium points seem to appear out of the blue, the **Transcritical bifurcation**, where the stability of two lines in the bifurcation plot alternate their stability, and the **Pitchfork bifurcation**, where one line of equilibrium points splits into three new lines. Check for yourself how these bifurcation points come to stand. More information on this can be found in Strogatz [2], but familiarizing yourself with the shape of these bifurcations will provide for the assignment.

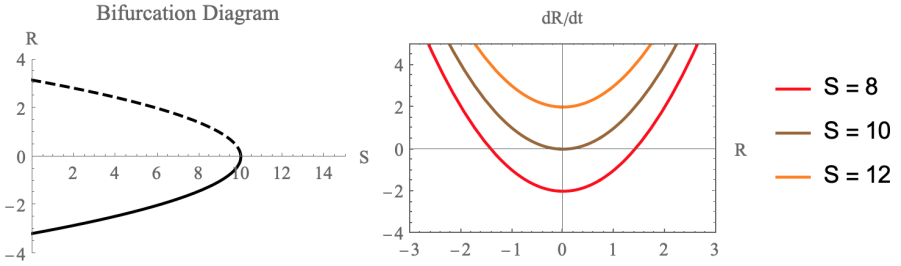

Figure 5: Saddle-Node Bifurcation

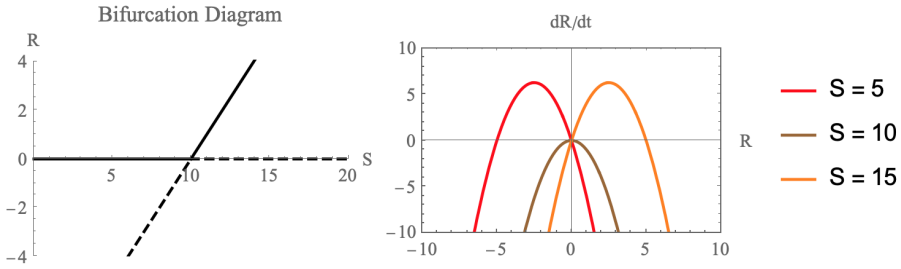

Figure 6: Transcritical Bifurcation

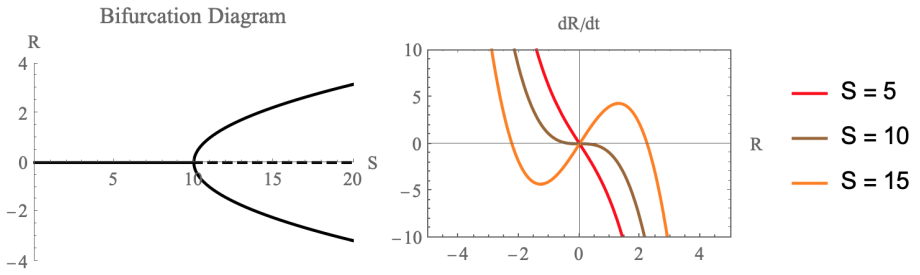

Figure 7: Pitchfork Bifurcation

Analysing the dynamics of two-dimensional systems

So far we have provided you with tools to analyse a one-dimensional dynamic system, by which we mean that the system only contains one dynamic variable. Now we will analyse an example of two-dimensional system, which is described as the set of two ODEs as shown below.

$$\begin{cases} \frac{da}{dt} = f(a,b) = a + b - 5 \\ \frac{db}{dt} = g(a,b) = 2a - 8b \end{cases}$$

We start by creating a **streamplot** (Figure 9), which displayes the flows of all the **trajectories** of the system. These **trajectories** are the solutions of the system, starting from different initial values. All trajectories can be found in the **phase plane**, in this case the (a,b)-plane. In these streamplots we can add **nullclines**, lines on which the rate equation is equal to zero. For the nullcline of $\frac{da}{dt} = 0$, this would mean that all trajectories crossing it only move in a vertical direction and for the nullcline of $\frac{db}{dt} = 0$ only in a horizontal direction. These nullclines help us getting an intuitive feeling on the dynamics in the phase plane and quickly indicate the equilibrium points on their intersections.

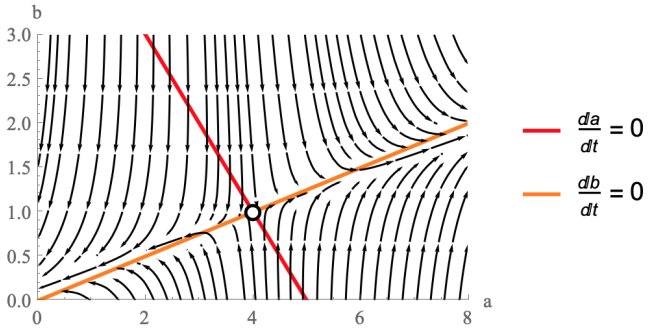

Figure 8: Streamplot with nullclines

Sometimes the stability of these equilibrium points is very clear from the vector plot, but it can also be a bit tricky to see. That is why determining the stability is the most reliable using the **Jacobian matrix**. The eigenvalues of the Jacobian matrix at the equilibrium points ($\{\lambda_1, \lambda_2\}$) provide information on the dynamics around these points. From these dynamics we can identify six different kinds of equilibrium points, as shown in Figure 9. A node has two real eigenvalues which are either both positive (**unstable node** (9a)) or both negative (**stable node** (9b)). A **saddle** (9c) has two real eigenvalues with opposite signs, so one is positive and one is negative, and is always unstable. When the eigenvalues at the equilibrium point are pure imaginary it is called a **center** (9d), an equilibrium point which is neither stable or unstable. The eigenvalues of a spiral are complex conjugates, where the real parts determine the stability. A positive real part results in an **unstable spiral** (9e) and negative real part indicates a **stable spiral** (9f).

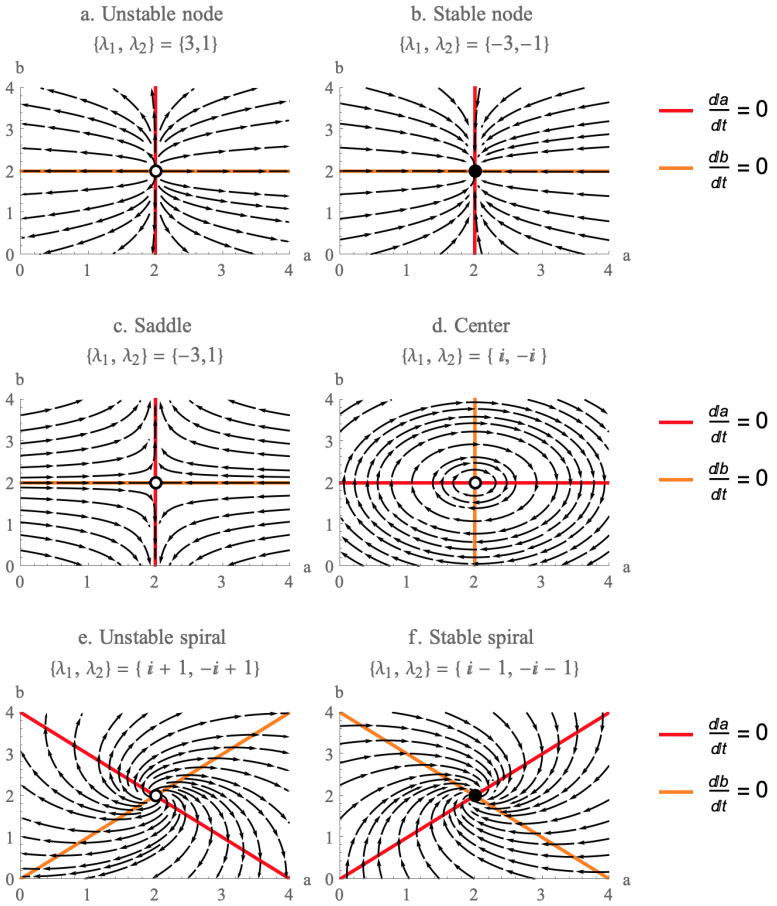

Figure 9: Streamplots with nullclines for different kinds of equilibrium points

Question 1: Perfect adaption

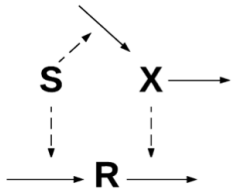

Figure 10: Network graphs perfect adaption

Figure 10 shows a network for perfect adaption. This behaviour is typical of chemotactic systems that respond to an abrupt change in attractants or repellents and then adapt to a constant level of signal. Our sense of smell operates this way and therefore this type of response is also called a "Sniffer". The kinetic equations for this system are given below.

$$\begin{cases} \frac{dR}{dt} = k_1 S - k_2 X R \\ \frac{dX}{dt} = k_3 S - k_4 X \end{cases}$$

$$k_1 = k_2 = 2, k_3 = k_4 = 1$$

(a) Plot the rate-equation of R for different values of S. (Hint: Remember that we use the steady-state solution.)

(b) Create the signal-response curve of this system and analyse its stability using (a). What is surprising about the signal-response curve?

(c) Plot the concentration of all species over time (t = 0 - 20) for $\{R_0, X_0\} = \{0, 0\}$. Let S increase by 1 every period of 4, starting at zero. How does your result relate to the signal-response curve?

Question 2: Feedback loops

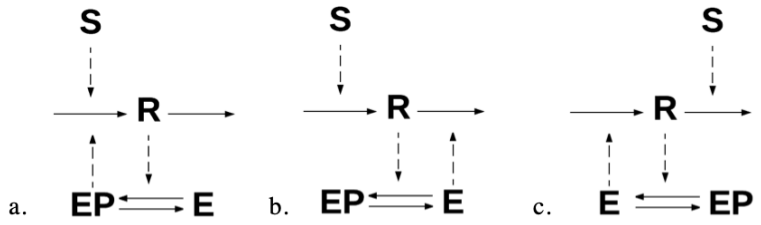

Figure 11: Network graphs of examples of feedback

In figure 11 the signal influences the response via two pathways that push the response in opposite directions. This is an example also of feed-forward control. An alternative would be that the response pathway feeds back to the signal, which can be either positive, negative or mixed. Figure 11 shows two types of positive feedback and one negative feedback example. In the **negative feedback**, the response R counteracts the effect of the stimulus S. In one type of positive feedback, called **mutual inhibition**, the signal S enhances the synthesis of the response, which in turn inhibits E by phosphorylation. In this case E degrades R and therefore E and R are mutually antagonistic. In another type of positive feedback called **mutual activation**, the signal S enhances the synthesis of the response, which in turn enhances its own synthesis via phosphorylating E. In the case of this mutual activation example, the feedback can create a discontinuous switch. This means that the cellular response changes abruptly and irreversibly as the signal magnitude crosses a critical value S_{crit}. Between the values of $S = 0$ and S_{crit}, the system is called bistable. This is also called the one-way switch and they play major roles in developmental processes and cellular apoptosis, both characterized by a point-of-no-return. In the case of the mutual inhibition example, the feedback can create a toggle-switch. When S is decreased enough, the switch will go back to the off-state. Between the values $S_{crit,1} < S < S_{crit,2}$, the response of the system can be small or large. The example of the negative feedback will cause a steady state concentration of the response R confined to a narrow window for a broad range of signal strengths.

Negative feedback: Homeostasis

$$\begin{aligned}\frac{dR}{dt} &= k_0 E(R) - k_2 S R \\ E(R) &= G(k_3, k_4 R, J_3, J_4) \\ k_0 &= 1, k_2 = 1, k_3 = 0.5, k_4 = 1, J_3 = J_4 = 0.01 \\ S &= 0.5, 1, 1.5\end{aligned}$$

Positive feedback: Mutual inhibition

$$\begin{aligned}\frac{dR}{dt} &= k_0 + k_1 S - k_2 R - k_2' E(R) R \\ E(R) &= G(k_3, k_4 R, J_3, J_4) \\ k_0 &= 0, k_1 = 1, k_2 = 0.05, k_2' = 0.5, k_3 = 1, k_4 = 0.2, J_3 = J_4 = 0.05 \\ S &= 1, 1.5, 2\end{aligned}$$

Positive Feedback Mutual activation

$$\begin{aligned}\frac{dR}{dt} &= k_0 E(R) + k_1 S - k_2 R \\ E(R) &= G(k_3 R, k_4, J_3, J_4) \\ k_0 &= 0.4, k_1 = 0.01, k_2 = k_3 = 1, k_4 = 0.2, J_3 = J_4 = 0.05 \\ S &= 0, 8, 16\end{aligned}$$

G refers to the Goldbeter-Koshland function:

$$G(u, v, J, K) = \frac{2uK}{v - u + vJ + uK + \sqrt{(v - u + vJ + uK)^2 - 4(v - u)uK}}$$

(a) Plot the rate equations of R for the different values of S.

(b) Create the signal-response curves of the networks and analyse its stability using (a). Also look for bifurcation points and classify them.

(c) Determine which network corresponds to which feedback type using (b).

(d) Describe your result on a molecular level.

(e) In the homeostatic system, explain why the response values are confined to a narrow window.

Question 3: The LAC-operon

An operon is a part of the DNA containing a cluster of genes controlled by a single promoter. If we take a closer look at the Lac-operon of Escherichia coli in Figure 12, we see the promoter which is repressed by a repressor in the absence of lactose. If lactose is available in the medium, import and conversion of lactose into allolactose occurs. Allolactose (A) is able to inhibit the inhibitor of the promoter and thereby allowing RNA polymerase to bind to the promoter and subsequently produce mRNA (M). The transcript contains information for the production of three enzymes: permease, β-galctosidase and transacetylase. The first two are respectively involved in cellular import of lactose and its enzymatic degradation, this mechanism is utilized by E. coli to derive energy from lactose.

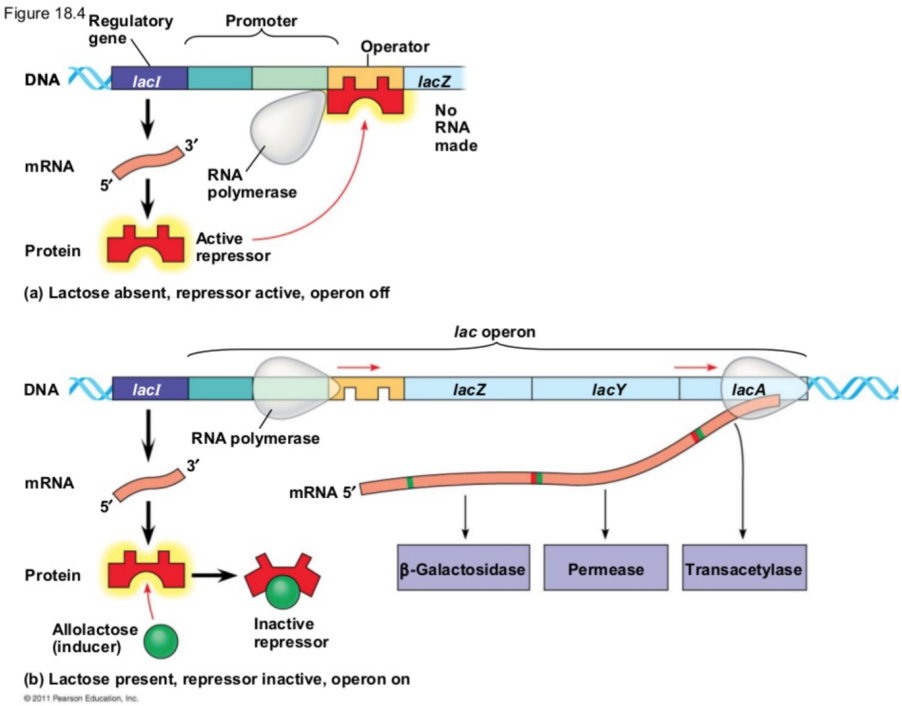

Figure 12: Schematic of the lac operon. The top scheme (a) shows the active respressor binding the operator in the absense of (allo)lactose, therefore inhibiting the transcription of the lac genes. Bottom scheme (b) shows deactivation of the repressor by allolactose, thereby deblocking the operator opening the way for the transcription of the lac genes.

Note that we make the very important assumption that the production of mRNA is proportional to its translation into the proteins, which allows us to write the following system of ODEs [3].

$$\begin{cases} \frac{dM}{dt} = k_1 + k_2(1 - \frac{1}{1+A^n}) - k_3 M \\ \frac{dA}{dt} = k_4 ML - k_5 A - V_{max} \frac{MA}{K_m + A} \end{cases}$$

(a) Interpret all of the terms in the two differential equations and describe both the positive and negative feedback loop.

(b) Create streamplots for $k_1 = 0.05, k_2 = k_3 = V_{max} = k_4 = 1, k_5 = 0.2, K_m = 2, n = 5$ for L = 1,2 and 3. Also include the nullclines.

(c) Obtain the equilibrium-points and identify them using the Jacobian-matrix. Do these equilibrium points represent dynamic or thermodynamic equilibriums? Explain your answer. (Hint: It might be convenient to derive the Jacobian matrix by hand)

(d) Plot A and M over time (t = 0 - 90) for $\{M_0, A_0\} = \{0, 0.05\}$. Start with L = 1 and increase L by 1 every period of 30. Describe your plot and elaborate on how it relates to the streamplots in (b). Does this system remind you of one of the previous networks?

(e) Elaborate on your results on a molecular level.

Question 4: Ionic Channels

An ionic channel in a cellular membrane (Fig. 13) can be regarded as special enzymes for mediating ion transportation over a membrane. They can exist in three states : S_1, S_2, S_3. S_1 and S_3 represent closed states on either side of the membrane (no ions can pass through the channel) and S_2 is an open state.

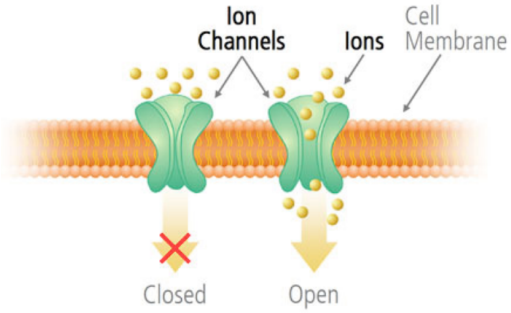

Figure 13: An ion channel in open and closed state [4]

On a given cell there are a number of channels. The fraction of channels in the states S_1, S_2 and S_3 are designated x, y and z respectively. Transitions between the states follow the schematic diagram below, where k_1, k_2 and k_3 are positive rate constants.

$$S_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} S_2 \xrightarrow{k_3} S_3$$

$$\begin{cases} \frac{dx}{dt} = k_2 y - k_1 x \\ \frac{dy}{dt} = k_1 x - (k_2 + k_3) y \\ \frac{dz}{dt} = k_3 y \end{cases} \qquad k_1, k_2, k_3 > 0$$

(a) Does this system go to an equilibrium? If so, try to describe it. Explain your answer using the schematic diagram and the set of ODEs. (Hint: Use that all rate constants are strictly positive)

(b) Create and analyze the streamplot of the x,y-plane for $\{k_1, k_2, k_3\} = \{2, 2, 1\}$. Do your conclusions correspond with your predictions in (a)?

(c) Plot x, y and z over time for the initial values of $\{x_0, y_0, z_0\} = \{1, 0, 0\}$ and $\{x_0, y_0, z_0\} = \{0, 1, 0\}$. Does your result correspond to your predictions in (a) and (b)?

Now assume $k_1 = 0$.

(d) How does this change the system on a molecular level?

(e) Rewrite the ODEs and analyze them, does the system still go to an equilibrium? If so, try to describe it.

(f) Plot x, y and z over time for the initial values of $\{x_0, y_0, z_0\} = \{0, 1, 0\}$ and $\{k_2, k_3\} = \{2, 1\}$. Do your conclusions correspond with your predictions in (e)?

Now we will analyse the values of our equilibrium points of x, y and z.

(g) Explain why we can define the equilibrium points as in the following equation.

$$\lim_{t \to \infty} \{x(t), y(t), z(t)\} = \{x^*, y^*, z^*\}$$

Now we will try to derive the values of $\{x^*, y^*, z^*\}$ from our set of ODEs we obtained in (e) and the initial values as used in (f). We start by obtaining the function of y(t) by integrating $\frac{dy}{dt}$ as shown below. Check the derivation for yourself.

$$\begin{aligned}
\frac{dy}{dt} &= -(k_2 + k_3)y \\
\frac{1}{y}dy &= -(k_2 + k_3)dt \\
\int_{y(0)=1}^{y(t_1)} \frac{1}{y}dy &= -(k_2 + k_3)\int_0^{t_1} dt \\
\big[ln(y)\big]_1^{y(t_1)} &= -(k_2 + k_3)\big[t\big]_0^{t_1} \\
ln(y(t_1)) - ln(1) &= -(k_2 + k_3)(t_1 - 0) \\
y(t_1) &= e^{-(k_2+k_3)t_1} \\
y(t) &= e^{-(k_2+k_3)t}
\end{aligned}$$

(h) Use $y(t)$ to show that $y^* = 0$. (Hint: take $t \to \infty$)

(i) Obtain the values of x^* and z^* by completing the following derivations on the dotted line. You may use that $\int_{t_1}^{t_2} y(t)dt = -\frac{1}{k_2+k_3}\big[y(t)\big]_{t_1}^{t_2}$. (This should take only a few steps.)

$$\frac{dx}{dt} = k_2 y$$
$$dx = k_2 y(t) dt$$
$$\int_{x(0)=0}^{x^*} dx = k_2 \int_0^{\infty} y(t) dt$$
$$\ldots$$
$$x^* = \frac{k_2}{k_2 + k_3}$$

$$\frac{dz}{dt} = k_3 y$$
$$dz = k_3 y(t) dt$$
$$\int_{z(0)=0}^{z^*} dz = k_3 \int_0^{\infty} y(t) dt$$
$$\ldots$$
$$z^* = \frac{k_3}{k_2 + k_3}$$

(j) Do your analytically obtained equilibrium points correspond to your equilibrium points in (f)?

References

[1] Le Novère, N., 2015. Quantitative and logic modelling of molecular and gene net- works. Nature Reviews Genetics.

[2] Strogatz, S. H., Nonlinear Dynamics and Chaos. 2nd edition, 2015

[3] De Boer, R.J. Ten Tusscher, k. Theoretical Biology. http://theory.bio.uu.nl/rdb/book- s/tb.pdf

[4] http://blog.canacad.ac.jp/wpmu/15matser/2011/09/14/4-1-passive-transport/