Chapter 1

Dynamics

1.1 Introduction

An important part of systems biology is the concept of modeling the dynamics of biochemical networks where molecules are the nodes and the molecular interactions are the edges. Due to the size and complexity of such a network, intuition alone is not sufficient to fully grasp its dynamical behavior. Instead a mathematical description of the network dynamics is used, which allows for testing and predicting the behavior in computer simulations.

This text starts with an introduction to dynamical systems. It then describes building-blocks which are used when modeling molecular interactions, and introduces how these are combined to model large biochemical networks. In the end there is a discussion on diffusion and the combination of reaction and diffusion into one model. The text is very sparse and are thought of as lecture notes for both teacher(!) and students.

Aim

The main purpose for this part of the course is to

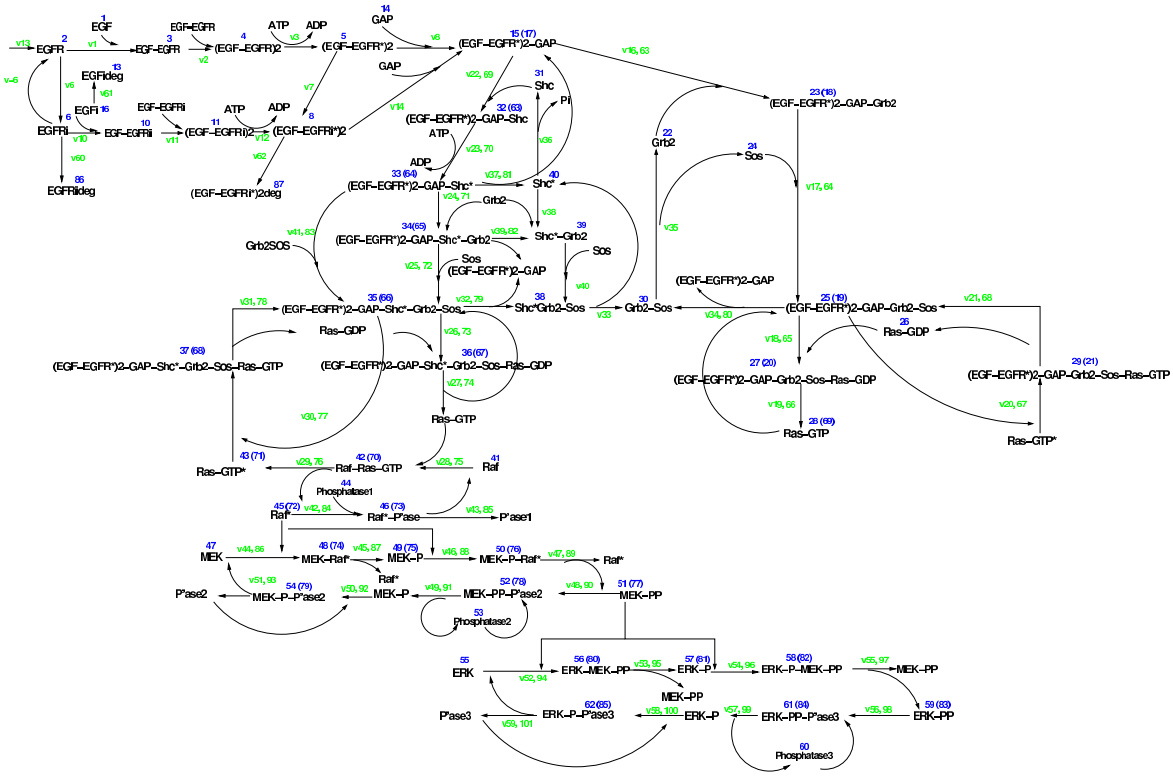
1. Understand the concept of modeling dynamical systems. Be able to create a mathematical model of a dynamical system and to do simple analysis of behavior and sensitivity.

2. Learn about some basic building-blocks for describing biochemical interactions (reactions,transcription,...). Be able to formulate these interactions in an ordinary differential equation framework.

3. Use the building blocks to create models of a complete biochemical network. Be able to do simulations of such a network in the computer excercise.

4. Understand the concept of diffusion and why it can be important when creating models in systems biology. Be introduced to reaction-diffusion models.

A main part for modeling in systems biology is items 2 and 3, which accordingly will constitute the bulk of the course. A main goal is to understand how a model of a large-scale network (as in the figure below) is developed.



Literature

This document is the lecture notes for the dynamics part of the systems biology course, and it is also the course literature. Additional suggested literature and articles will be availabe in pdf-format at http://www.thep.lu.se/~henrik/bnf079/literature.html. The compilation consists of a number of introductory texts and scientific publications, and can be used as references for the interested reader to clarify concepts and to learn more about specific examples. It is not mandatory to read these extra references to be able to pass the course.

Contact

Henrik Jönsson, henrik@thep.lu.se, Ph 046-2220667.

1.2 Model building in systems biology

Before building a mathematical model of a biological system, it is important to do some basic decisions on how the model should be defined. Some examples of these decisions are:

- **Resolution.** What should be the resolution of the model? What are our model variables representing? Given the ample experiments of quantitative molecular data that has been developed during the past years, we can now build models describing molecular contents (numbers or concentrations) and compare directly to experiments. In this course we will look at dynamical models of molecular contents describing biochemical networks within single cells and touch upon multicellular systems. The choise of this resolution would for example not be applicable to explain the evolution of the human population (for which it would be far too detailed) or describe a single molecular reaction mechanism in detail (for which the resolution is too low and an quantum mechanical approach would be needed).

- **Continuous vs. discrete.** Molecules are individual objects, and the quantitative measure of molecular content is in principle number of molecules. On the other hand, the number of molecules (of the same type) within for example a cell is most often large and a continuous variable for the concentration (number of variables per unit volume) is then applicable to describe the system behavior. In this course we will use continuous concentrations as variables. The limit of when the concentration is sufficient to describe a system do of course depend on the details of the system but typically when the number of molecules are more than $10^1 - 10^2$ it is safe to use concentration as a measure of molecular content.

- **Deterministic vs. stochastic.** This point is somewhat related to the previous point. In principle there is a probability connected to an individual reaction to occur, which can be taken into account using a stochastic update of the system variables (reactions happen with a specific probability). Again, within this course we will assume systems with a large number of molecules where it most often is applicable to use a deterministic description of the system update.

A modeling approach for a biochemical network includes several different steps of tasks to be solved. For the work to be effective biological experiments have to be combined with the theoretical modeling work. Important steps whithin a modeling approach are

1. **Define the molecular players and interactions.** It is a slow and hard experimental work to be able to define which molecules that are involved in different biological processes, and how these interact. The genome projects, where the complete genome of different species are sequenced, have increased the knowledge about the molecular (protein) players. This provides a list of components, and molecular genomics research are contributing to the knowledge of what biological processes

individual molecules are involved in, and how these interact. The result is biological network which you have come across earlier in the course. We will assume that this is the input to our modeling approach and will not discuss this any further. It should be noted though, that one main purpose of a modeling approach is to be able to guide these kinds of experiments for an increased understanding of the biological system at hand.

2. **Describe the molecules and interactions in a mathematical model.** To be able to do a quantitative model of a biochemical network, molecular concentrations have to be turned into variables and their interactions need to be described by mathematical functions. These functions do depend on the type of interactions that are described. A main goal of this part of the course is to learn what type of mathematics that is used to describe the interactions for e.g. molecular reactions and gene transcription, and how these individual interactions are combined into a complete quantitative model for the network.

3. **Estimate parameter values for the model.** The mathematical description includes a number of parameters defining reaction rates etc. Different values of these parameters can result in completely different behavior of the model. Hence it is crucial to estimate the parameter valus that are relevant for the specific biological network. One way to do this is to *experimentally measure* the parameter for a specific reaction. The benefit of this is that it will result in a single value for the parameter. A drawback is that it is hard to do such measurements within a biological organism, and if it is measured elsewhere that specific condition might lead to a different value compared to within the organism. Another approach is *reverse engineering*, where model parameters are estimated by fitting model output to available experimental data. This will exclude most parameter values but still it is possible that this approach will find different values for a single parameter that equally well describe the biological behavior. Within this part of the course we will see how parameter estimations can be done in practice.

4. **Analyse the dynamical behavior.** A final step in a modeling approach is to analyse the behavior of the defined model. Many molecular networks and modules show very high robustness. This should then also be accounted for in the model and can be tested by a sensitivity analysis, where the changes of behavior is tested when parameters are perturbed. Also, the model can be tested for perturbations where molecules or interactions are removed from the system, which then can be compared with knock-out experiments, or provide biological predictions from the model. The analysis can provide feed-back into the previous three steps improving the description and knowledge of the biological system.

1.3 Dynamics

Dynamics deals with changes; the evolution in time of a system. It can concern more or less anything from e.g. classical mechanics with an apple falling to the ground, or the growth of the human population. Within systems biology dynamics typically refer to the changes in molecular concentrations (or numbers) within a cell.

A system is defined by i) a set of variables defining the state of the system, and ii) the rules for how the variable values change in time. Variables can be discrete where the state of the variable can be described by a distinct set of values, or continuous where any real value is allowed. The update rules can depend on the time and on the state of all variables. It can be deterministic where the time and variable states uniquely defines the state at next time point, or it can be stochastic where the time and variable state defines the probability of how the variable values changes over time.

The goal when dealing with a dynamical system is to describe and analyse the behavior of the individual variables and also of the complete system. A dynamical system can be in equilibrium where variables do not change, it can oscillate in a repeating fashion, or it can be more complicated and even chaotic. We will only touch on these subjects briefly, and the interested student can learn more in introductory courses or text books of the subject (e.g. fys244, System theory, which is given by the Complex Systems group).

1.3.1 Ordinary differential equations

A fundamental concept when studying dynamics for a continuous system is ordinary differential equations (ODEs). Within this course, we will only deal with systems defined as

$$\begin{aligned} \frac{dx}{dt} &= f_x(x, y, ..., t) \\ \frac{dy}{dt} &= f_y(x, y, ..., t) \\ &... \end{aligned} \tag{1.1}$$

$x, y, ...$ are the state variables which in our case typically are molecular concentrations, and $f_x, f_y, ...$ are the functions describing the molecular interactions. The dimension of a system is defined by the number of variables that are included. If the differential equations are given and initial state (values) of the variables are known, the future behavior of the system is completely defined.

Numerical integration of ODEs

The systems of ODEs for molecular networks are most often too complex to solve analytically and numerical integration is used to simulate the behavior on a computer. There are many sophisticated algorithms for doing this, but almost all are built from discretizing the differential equation and step forward in time with small steps. The simplest

variant of this stepping is the Euler step

$$\begin{aligned} \frac{\Delta x}{\Delta t} &= \frac{x(t+\Delta t)-x(t)}{\Delta t} = f_x(x,y,...,t), \quad \text{or} \\ x(t+\Delta t) &= x(t)+\Delta t f_x(x,y,...,t). \end{aligned}$$

The error introduced by this step is of the order Δt^2 at each step. More accurate solvers will be discussed within a computer exercise.

1.3.2 Behavior of a dynamical system

For one dimensional systems the only possible behavior is that the variable value approaches a specific value, which is defined as a fixed point. The variable might also approach plus (or minus) infinity.

Example: creation and degradation of a molecule

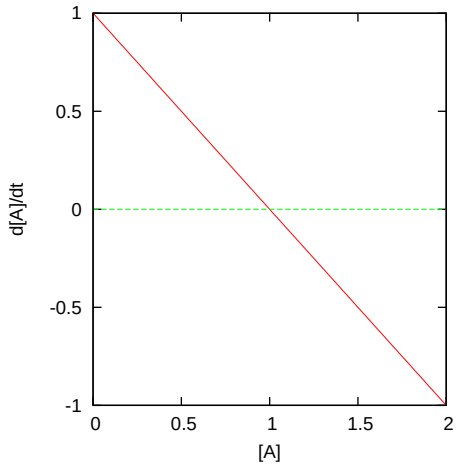
Assume a molecule A which is produced and degraded at a constant rate.

$$\emptyset \underset{d}{\overset{k}{\rightleftharpoons}} A, \tag{1.2}$$

where k is the production rate and d is the degradation rate. The production is assumed to be constant in time (or depend on variables that do not change in time and hence are left outside the model). The degradation rate is assumed to be constant for each induvidual molecule of A. A differential equation describing this system is

$$\frac{d[A]}{dt} = k - d[A], \tag{1.3}$$

where $[A]$ is the concentration of molecule A. Fixed points of the system can be found by solving the algebraic equation $d[A]/dt = 0$ (i.e. if the system is in such a state it will stay in this state). k and d are assumed to be positive constants, and the only solution is $[A] = k/d$. A closer look at the time derivative as a function of the concentration of $[A]$ (see figure) resolves more of the dynamical behavior.



Since the derivative is positive when $[A] < k/d$ and negative when $[A] > k/d$ the system will always approach k/d for infinite times. Since any initial concentration $A(t_0)$ eventually will lead to the fixed point, it is called globally stable.□

Example: autocatalysis

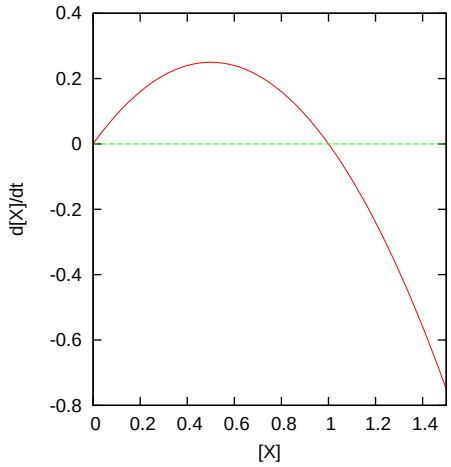
In this example there is a molecule X which induces it's own production mediated by a molecule A.

$$A + X \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 2X, \tag{1.4}$$

The law of mass action (which will be discussed in more detail later) states that the rate of a reaction is proportional to the concentrations of the reactants. In this case we assume that there is a surplus of molecule A resulting in that its concentration can be assumed to be constant.

$$\frac{d[X]}{dt} = k_1[A][X] - k_2[X]^2 = K[X] - k_2[X]^2 = [X](K - k_2[X]), \tag{1.5}$$

where the constant $K = k_1[A]$ is introduced. $d[X]/dt$ equals zero for either $[X] = 0$ or $[X] = K/k_2$. A closer look at $d[X]/dt$ as a function of $[X]$ reveals that the fixed points are of different kinds. The $[X] = 0$ fixed point is instable, while $[X] = K/k_2$ is stable (Figure).



The conclusion is that only if we start the system exactly at $[X] = 0$ it will stay there. For any other initial value, the system ends up in $[X] = K/k_2$. A quite interesting note to make is that the equation in this example is exactly the logistic equation used in population dynamics.□

The two examples show that it is possible to analyse the behavior of of a dynamical system without solving the differential equation. We can completely predict what will happen in those examples.

For one dimensional systems we can formalize the approach. Given a differential equation

$$\frac{dx}{dt} = f(x) \tag{1.6}$$

1. Find all fixed points x^* by solving $dx/dt = 0$.
2. Investigate the sign of dx/dt around each fixed point to determine the stability. This can be done by plotting it as in the examples, but also a little bit more formal by looking at $df(x)/dx$ in the fixed points, where

   $\frac{df(x^*)}{dx} < 0 \rightarrow$ Fixed point stable

   $\frac{df(x^*)}{dx} > 0 \rightarrow$ Fixed point instable

   while if the derivative is zero at the fixed point further analysis is needed (it is typically semistable).

It is quite important to note that the behavior (and analysis) do depend on the values of the parameters. Different parameter values can result in different stabilities (e.g. a change from stable to unstable). What will for example happen if d in our first example is negative?

Systems of higher dimensionality can have more elaborate behaviors such as oscillations and chaotic behavior. The analysis of higher dimensional systems are out of scope for this course, and we will only study an example.

Example: two autocatalysing molecules that form a complex

This example is an extension of the previous example, where we now have two molecules X, Y which induces their own production mediated by molecules A and B. X and Y can also form a complex C $(= XY)$.

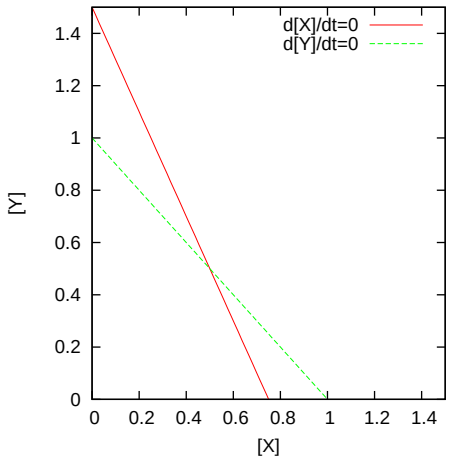
$$\begin{aligned} A + X &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} 2X, \\ B + Y &\underset{k_5}{\overset{k_4}{\rightleftharpoons}} 2Y, \\ X + Y &\overset{k_3}{\rightarrow} C, \end{aligned}$$

We again assume that there is a surplus of A and B, resulting in their concentrations being constant. Since the dynamics of X and Y does not depend on the complex C, it will also be left out of the analysis.

$$\begin{aligned} \frac{d[X]}{dt} &= k_1[A][X] - k_2[X]^2 - k_3[X][Y] = K_1[X] - k_2[X]^2 - k_3[X][Y] \\ &= [X](K_1 - k_2[X] - k_3[Y]), \\ \frac{d[Y]}{dt} &= k_4[B][Y] - k_5[Y]^2 - k_3[X][Y] = K_2[Y] - k_5[Y]^2 = -k_3[X][Y] \\ &= [Y](K_2 - k_5[Y] - k_3[X]), \end{aligned}$$

where the constants $K_1 = k_1[A]$ and $K_2 = k_4[B]$ are introduced. $d[X]/dt$ equals zero for either $[X] = 0$ or $[X] = (K_1 - k_3[Y])/k_2$. These expressions no longer defines specific points but rather defines lines which are defining nullclines. The nullclines for Y are similarly defined by $[Y] = 0$ and $[Y] = (K_2 - k_3[X])/k_5$.

An informative way of representing this system is by plotting the nullclines in the phase space (Figure) which is a plot where $[X]$ and $[Y]$ defines the axes. Now it is easy to see that for example the $[X] = 0$ null cline corresponds to all points on the $[Y]$ axis. Fixed points of the system are found where the nullclines intersect where both $d[X]/dt$ and $d[X]/dt$ are zero. In the regions in between the nullclines there are non-zero time derivatives (for both $[X]$ and $[Y]$) and by looking at the signs of the derivatives it is possible to analyse the dynamics. It can for example be seen that $d[X]/dt$ is positive beneath the nullcline defined by $[X] = (K_1 - k_3[Y])/k_2$ and negative above.



The conclusion of this analysis is that there are four fixed points $(0,0)$,$(K_1/k_2,0)$,$(0,K_2/k_5)$, and '$([X]^*,[Y]^*)$'. The only stable fixed point is at '$((K_1k_5-K_2k_2)/(k_3(k_5-k_2)),[Y]^*)$'.

Again it must be noted that this is for the parameter set used, and the behavior can change if for example the nullclines for $[X]$ and $[Y]$ overlaps (the two not defined by the axes).□

1.3.3 Sensitivity analysis of a dynamical system

Biological systems have evolved and survived for millions of years. They typically inherit a stability towards fluctuations in parameters, and when creating a model this should also be accounted for. A simple measure for sensitivity is to measure the relative change of a system feature due to a change in a parameter. For example the feature can be the equilibrium concentration of a compound, C for which the sensitivity (S) to a parameter p is

$$S_p = \frac{\frac{dC}{C}}{\frac{dp}{p}} = \frac{dC}{dp}\frac{p}{C} \approx \frac{\Delta C}{\Delta p}\frac{p}{C} \tag{1.7}$$

Example: creation and degradation revisited

Let's go back to our first example where a molecule A is produced and degraded at constant rates.

$$\emptyset \underset{d}{\overset{k}{\rightleftharpoons}} A, \tag{1.8}$$

where k is the production rate and d is the degradation rate. We calculated that this system had a fixed point for $A^* = k/d$. This system is so simple that it is possible

to calculate the sensitivity of the fixed point with regard to the two parameters. The derivative form leads to

$$\begin{aligned} \frac{dA^*}{dk}\frac{k}{A^*} &= \frac{1}{d}\frac{kd}{k} = 1 \\ \frac{dA^*}{dd}\frac{d}{A^*} &= -\frac{k}{d^2}\frac{dd}{k} = -1 \end{aligned} \tag{1.9}$$

The difference version relies on that a parameter value is changed with a fraction f ($p \to p + fp$), and that the fixed point is calculated (or measured in a simulation) for the new parameter value. Changing the parameters a fraction f leads to new fixed points

$$\begin{aligned} A^*(k + fk, d) &= \frac{k + fk}{d} = (1 + f)\frac{k}{d} \\ A^*(k, d + fd) &= \frac{k}{d + fd} = \frac{1}{(1 + f)}\frac{k}{d} \end{aligned} \tag{1.10}$$

and the sensitivity measures are given by

$$\begin{aligned} \frac{A^*(k + fk, d) - A^*(k, d)}{fk}\frac{k}{A^*(k, d)} &= \frac{fA^*(k, d)}{fk}\frac{k}{A^*(k, d)} = 1 \\ \frac{A^*(k, d + fd) - A^*(k, d)}{fd}\frac{d}{A^*(k, d)} &= \frac{-\frac{f}{1+f}A^*(k, d)}{fd}\frac{d}{A^*(k, d)} = -\frac{1}{1 + f} \approx -1 \end{aligned} \tag{1.11}$$

where in the last equation f is assumed to be small.

We can see that if the two parameter parts are summed we get zero (summation law), and that when using the difference version f needs to be small not to introduce errors. The conclusion is that the fixed point is directly increased with the same fraction as k is changed. For the d parameter there is an decrease of the same fraction as d is varied. The system is sensitive to changes in the parameters which is obvious since the parameters are determining the dynamics (and the fixed point) directly.□

It should be noted that this sensitivity measure is local and depends on the current system "topology" and most often on parameter values. When applying a sensitivity measure, there are often summation laws appearing, as for example in the case of measuring sensitivity on equilibrium values $\sum_i S_{p_i} = 0$. Features often used in robustness (anti-sensitivity) analysis are e.g. the time integral of a variable, the duration or amplitude of a peak, etc.

1.4 Biochemical rate equations

In a deterministic continuous formulation, molecular reactions are described by differential equations defining the rate of change in molecular concentrations. Molecular concentrations are most often measured in molar which is defined by mole per liter, where one mole is 6.02×10^{23} molecules. Typical molecular concentrations within a cell are from $0.1nM$ to $1\mu M$ (with lots of exceptions of course).

1.4.1 Mass action formalism

Despite its simplicity, the mass action formalis has been validated in many experimental settings. The law of mass action states that the rate of an elementary chemical reaction is proportional to the product of the concentrations of the reactants. It is based on the assumptions of i) a well stirred solution and ii) low molecular concentrations, where the probability of diffusing molecules to get close enough, for a reaction to occur, is proportional to the concentrations. A rate parameter is used to define the 'probability' of a reaction to occur if two molecules approach each other.

Generally a mass action reaction can be written as

$$s_1S_1 + s_2S_2 + ... \stackrel{k_f}{\rightarrow} p_1P_1 + p_2P_2 + ...,$$

where the varaibles $S_1, S_2, ...$ are defining the reactants and the $P_1, P_2, ...$ are defining the products. The parameters $s_1, s_2, ..., p_1, p_2, ...$ are called stoichometric coefficients and k_f is the rate parameter. The stoichometric coefficients is typically chosen such that the total mass is conserved in the reaction (or such that atom numbers are the same before and after the reaction).

Example: a simple mass action reaction

Consider the simple reaction of species A and B forming complex C.

$$A + B \underset{k_b}{\overset{k_f}{\rightleftharpoons}} C.$$

k_f is the rate of the forward reaction of unit $[time]^{-1}[conc]^{-1}$, while k_b is the rate of the backward reaction of unit $[time]^{-1}$. Note that reaction rates do not have a uniquely defined unit, but rather depends on the reaction. In a differential equation formalism the equations are defined by

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -\frac{d[C]}{dt} = -k_f[A][B] + k_b[C], \tag{1.12}$$

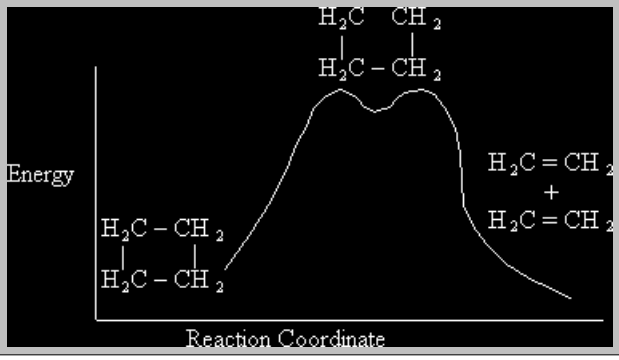
which will have an equilibrium point (fixed point) for $[C]/[A][B] = k_f/k_b$ where $K = k_f/k_b$ defines a relation between concentrations of reactants and products which is independent on initial concentrations. K is often defined as the reaction constant.□

1.4.2 Thermodynamics and rate constants

In experiments it can be seen that the logarithm of the rate constant, $\ln k$, is linearly related to the inverse temperature $1/T$. The parameters for the slope and intercept is formulated in Arrhenius law

$$k = Ae^{-E_a/TR} \tag{1.13}$$

where E_a is the activation energy, R is the gas constant and A is the steric factor, a constant measuring the efficiency of a molecular collision leading to a reaction.
In transition state theory the energy is replaced by the Gibbs free energy, $G = E + PV - TS$, where P is the pressure V is the volume, T is the temperature and S is the entropy. The idea is that the a molecule is in a local minima in a "reaction space", and that for a reaction to happen, it has to find a path to the product within this space, and a maxima needs to be passed (see figure below). Values for the Gibbs free energy for different molecules can be found in the literature and the reaction constant of a bidirectional reaction can be related to the difference in G.



1.4.3 Enzyme kinetics

Many reactions have a far too high activation energy to ever occur spontanously. A common type of reaction is an enzyme reaction, where a helper molecule (the enzyme) fascilitate a reaction to occur. The enzyme is not used up in the reaction itself.

Example: a simple enzymatic reaction

Consider the simple reaction of species A forming compound B with the help of enzyme E.

$$A + E \xrightarrow{k} B + E.$$

k is the rate of the reaction of unit $[time]^{-1}[conc]^{-1}$. Using a differential equation formalism the equations are defined by

$$\begin{aligned} \frac{d[A]}{dt} &= -\frac{d[B]}{dt} = -k[A][E], && (1.14)\\ \frac{d[E]}{dt} &= 0. && (1.15)\\ & && (1.16) \end{aligned}$$

The problem with this formulation is that there is no upper limit on how much a single enzyme molecule can facilitate the reaction. Often there is an upper limit on the rate due to the fact that the enzyme is occupied during the reaction, and a model accounting for this is described in the next section.□

1.4.4 Enzyme kinetics, Michaelis-Menten

A more proper description of an enzyme reaction is to have the enzyme E bind to the substrate S and having the substrate turn into a product P while the enzyme is released

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} SE \overset{k_3}{\rightarrow} P + E. \tag{1.17}$$

The rate equations for this system can be written as

$$\begin{aligned} \frac{d[S]}{dt} &= -k_1[S][E] + k_2[SE]\\ \frac{d[E]}{dt} &= -k_1[S][E] + k_2[SE] + k_3[SE]\\ \frac{d[SE]}{dt} &= k_1[S][E] - k_2[SE] - k_3[SE]\\ \frac{d[P]}{dt} &= k_3[SE] \end{aligned} \tag{1.18}$$

The first reaction is assumed to be fast (and in equilibrium) and we assume that $d[SE]/dt \approx 0$. Solving the fixed point equation gives $K = k_1/(k_2 + k_3) = [SE]/[S][E]$. If we also assume a constant amount of total enzyme, $[E] + [SE] = E_0$, the complex concentration can be written as a function of the substrate concentration,

$$\begin{aligned} [SE] = K[S][E] &= K[S](E_0 - [SE])\\ [SE]\,(1 + K[S]) &= KE_0[S]\\ [SE] = \frac{KE_0[S]}{1 + K[S]} &= \frac{E_0[S]}{(1/K + [S])}. \end{aligned} \tag{1.19}$$

The production of P as a function of the substrate concentration is then

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{1.20}$$

where the constants $V_{max} = k_3 E_0$ and $K_m = 1/K$. The choice of parameters is due to the fact that V_{max} is the saturated maximal rate of production and K_m is the amount of substrate that corresponds to half the maximal rate (Fig. 1.1). A problem with the

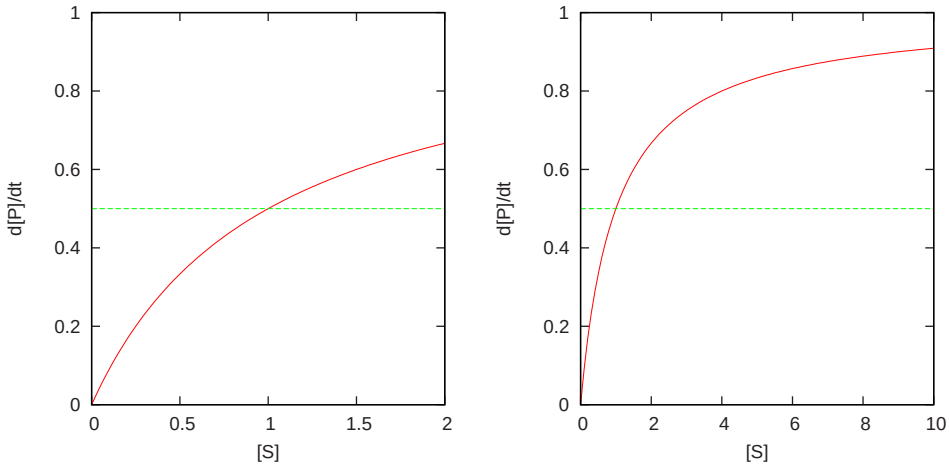


Figure 1.1:

Michaelis-Menten equation is the "slow" response to substrate concentration compared with what is often seen in experiments. To get the rate $0.1V_{max}$ a substrate concentration of $S_{0.1} = K_m/9$ is needed and to get a rate of $0.9V_{max}$, the substrate concentration needs to be $S_{0.9} = 9K_m$. Hence an 81-fold change in concentration is needed between 'on' and 'off' states. This is often handled by using a Hill-type kinetics as will be discussed in more detail later.

It should also be noted here that the dependence on the enzyme concentration is built into the V_{max} parameter and assumed to be constant. The amount of enzyme is often also a dynamic variable and the reaction is then described by

$$\frac{d[P]}{dt} = \frac{V'_{max}[S][E]}{K_m + [S]} \tag{1.21}$$

Example: protein activation/deactivation cycle

Previously in the course you have seen the example of a protein that can be activated and deactivated

$$X^* \underset{(V_1,K_1)}{\overset{(V_2E,K_2)}{\rightleftharpoons}} X,$$

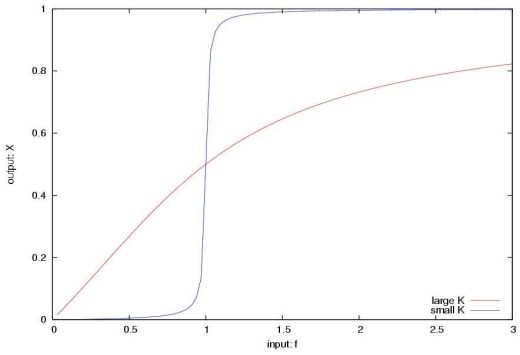
where the total concentration is constant $X^* + X = X_{\text{tot}} = 1$ (or $X^* = 1 - X$). Assuming that both the activation and deactivation are dependent on other molecules (enzymes), and that the activation enzyme is dynamic, result in the following Michaelis-Menten description

$$\frac{d[X]}{dt} = -\frac{V_1[X]}{K_1 + [X]} + \frac{V_2[E](1-[X])}{K_2 - (1-[X])}. \tag{1.22}$$

Setting the parameters $K_1 = K_2 = K$ and $f = V_2[E]/V_1$ and investigating the system at equilibrium ($\frac{d[X]}{dt} = 0$) results in the equation

$$\frac{[X]}{K + [X]} = f\frac{(1-[X])}{K + (1-[X])}. \tag{1.23}$$

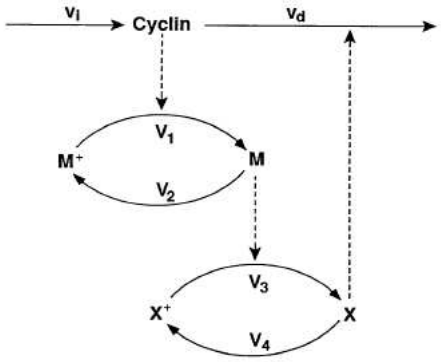
When studying how the activation, $[X]$, is dependent on the input, f, it was shown to behave either as an analogue amplifier or a digital switch depending on the K value, as shown in the figure.



□

Example: cell cycle

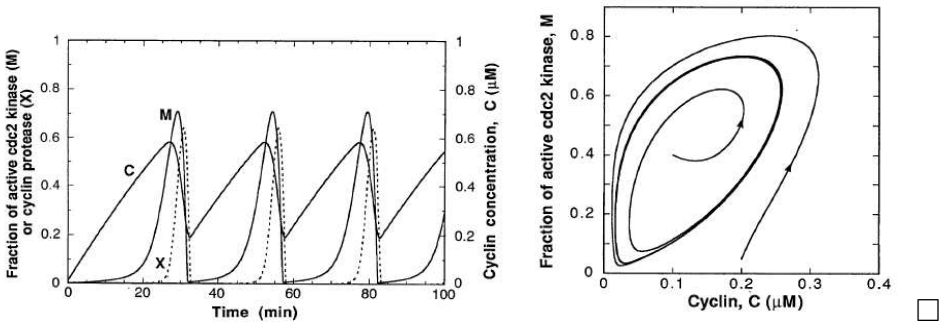
A minimalistic model for the cell cycle was introduced by Goldbeter 1991. It has only three state variables and the interactions are shown in the figure.



In the model, cyclin (C) is produced and degraded at constant rates. The cyclin induces a cyclin kinase (M) to be activated, which in turn activates a cyclin protease (X). Finally the protease induces degradation of the cyclin closing a feed-back loop in the system. All reaction kinetics used is in the Michaelis-Menten format. A minor simplification of the equations leads to the following model.

$$\begin{aligned}\frac{dC}{dt} &= v_i - v_d X\frac{C}{K_d + C} - k_d C \\ \frac{dM}{dt} &= V_1 C\frac{(1-M)}{K_1 + (1-M)} - V_2\frac{M}{K_2 + M} \\ \frac{dX}{dt} &= V_3 M\frac{(1-X)}{K_3 + (1-X)} - V_4\frac{X}{K_4 + X}\end{aligned} \tag{1.24}$$

Simulation of the network shows that, for some ranges of parameter values, an oscillatory solution is possible (which also exhibit limit cycle behavior) as can be seen in the figures below.



□

1.4.5 Models within a cell

The mathematical formulations described in previous sections are simplified and assumes idealized conditions. For example the assumptions of low molecular concentrations and of well-stirred solutions are very unlike the situation in a cell (Fig.1.2).

Example: generalized mass action

It is often the case that the mass action dynamics deviate from *in vivo* experiments. It might then be useful to "extend" the reaction models to better correlate with experiments. In the generalized mass action approach the concept of activity is introduced. The idea is that the effective concentrations for a reaction can be different from the absolute concentration. Without going into details, the generalized mass action formalism for a simple reaction

$$A + B \xrightarrow{k_f} C \tag{1.25}$$

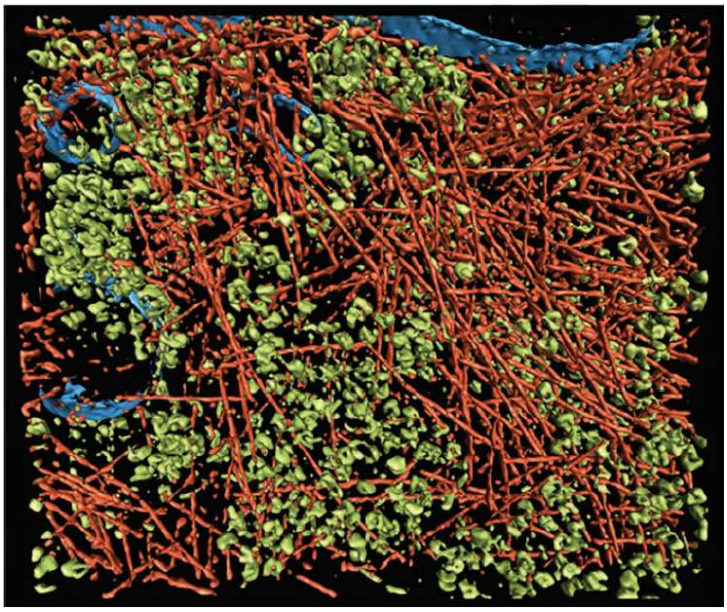


Figure 1.2: Visualization of actin network, membranes, and cytoplasmic macromolecular complexes in a volume of 815 nm by 870 nm by 97 nm. Colors were subjectively attributed to linear elements to mark the actin laments (reddish); other macromolecular complexes, mostly ribosomes (green); and membranes (blue). From Mendalia et. al. (2002), Science 298, 1209-1213. Copyright 2002 AAAS.

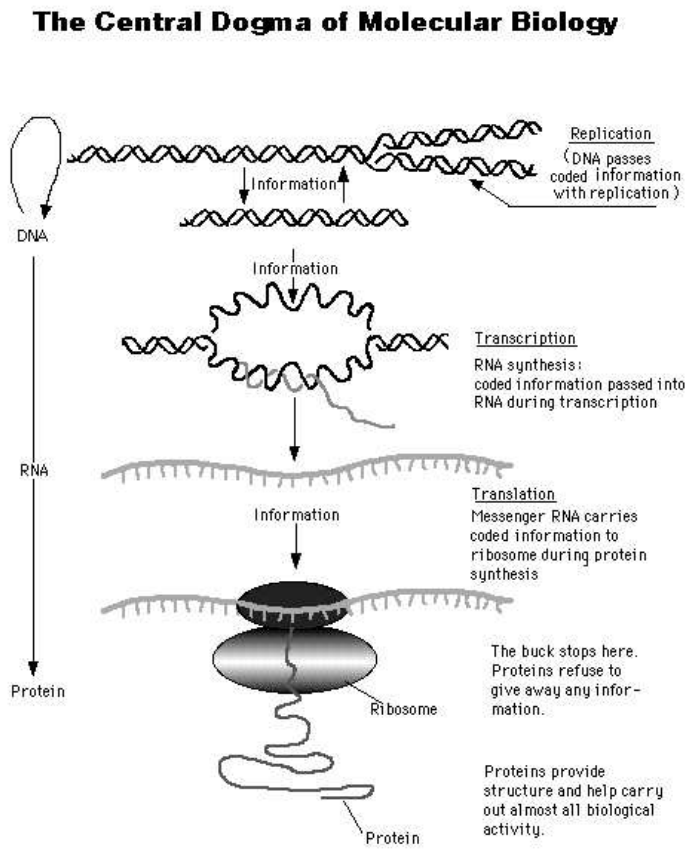
uses a differential equation of the form

$$\frac{d[A]}{dt} = k_f a[A]^\alpha b[B]^\beta \tag{1.26}$$

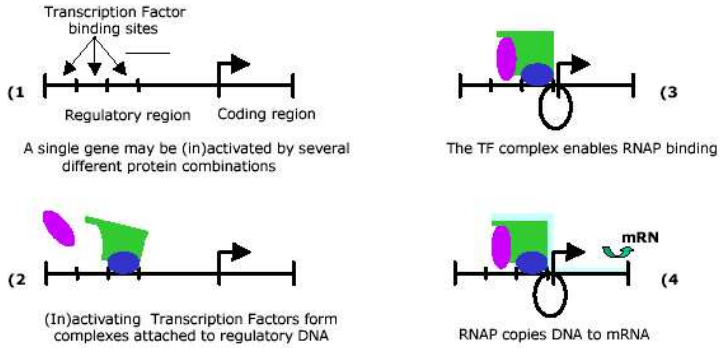
where a, α, b, and β are (real-valued) parameters. The generalized mass action hence allow for additional possibilities of dynamical behavior compared to classical mass action.□

1.5 Gene regulation

The central dogma of molecular biology concerns the information flow within cells. It states that the information is translated between different molecular types as follows:



For gene regulation the important steps are the transcription (DNA → RNA) and translation (RNA → Proteins). As have been discussed previously in the course, also the ability of specific proteins (transcription factors) to affect the transcription rate is essential (see figure below). This allows for a network of proteins affecting each others production (or a network of genes affecting each others activity).

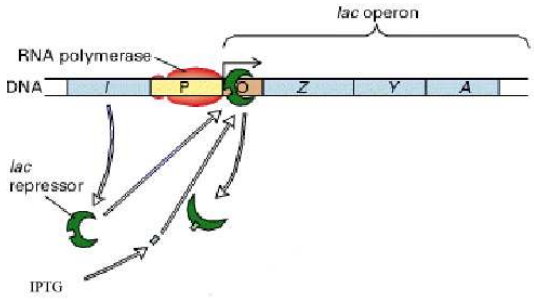


The biological processes involved in transcription and translation are overwhelmingly complex, and the mathematical descriptions we will discuss are simplified approxima-

tions. This allows for using them in large network settings, and is most often sufficient due to the lack of detailed experimental data. It is also often convenient to model transcription and translation within a single equation, which, due to the complex input-output relations, requires non-linear descriptions.

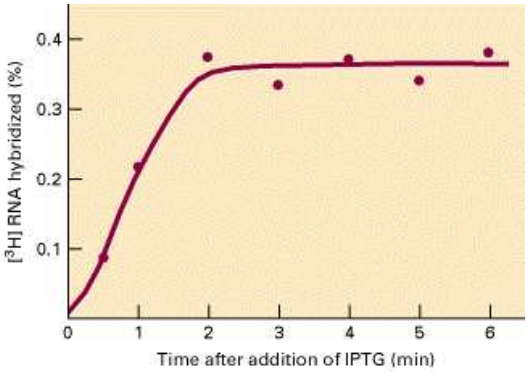
Example: the lac-operon

The idea that transcription factors (proteins) bind to the DNA and regulate the transcription rate of genes was first introduced by Jacob and Monod in 1961. They used the lac operon in *E. coli* and their model is shown in the figure.



In the model a transcription factor, lac-repressor, binds to the DNA and prevents transcription of the lac-operon. The repressor can form a complex with IPTG, which results in that the repressor is released from the DNA and transcription is activated.

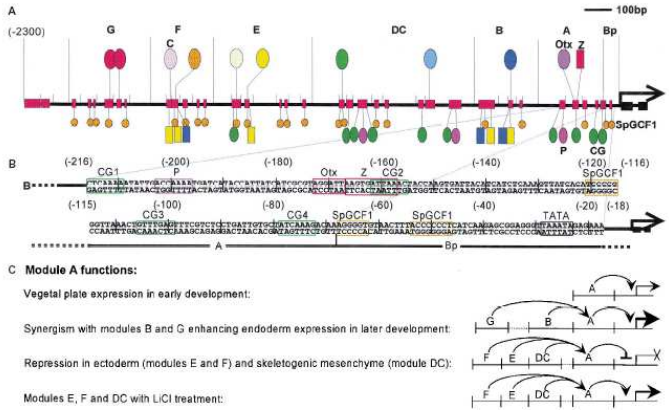
In an experiment where IPTG is introduced to the cells and the lac-operon activity is measured, a quick response can be seen (figure)



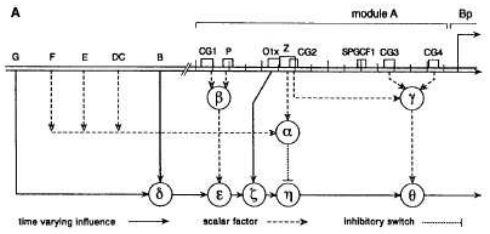
This simple gene regulation system has features which are common for gene expression. It is highly nonlinear, and it has a saturated behavior with a maximal value of the production rate.□

Example: sea urchin gene Endo16

When it comes to genetic regulation in multicellular organisms, one of the most studied species is the sea urchin. This example shows the complexity of a single promotor with a manifold of modules which in turn is regulated by a manifold of molecules (figure).



The authors have also created a model of the transcription activity and use a combination of logical rules and contiunous equations (figure below). Fortunatly(?), this complex regulation is beyond the scope of the course, but one should be aware of that the simple models introduced later in this section have limitations on how accurately they describe the transcription/translation processes.



B

if (F = 1 or E = 1 or CD = 1) and (Z = 1)	Repression functions of modules F, E, and DC mediated by Z site
α = 1	
else α = 0	
if (P = 1 and CG_n = 1)	Both P and CG_n needed for synergistic link with module B
β = 2	
else β = 0	
if (CG_3 = 1 and CG_4 = 1 and CG_n = 1)	Final step up of system output
γ = 2	
else γ = 1	
δ(t) = B(t) + G(t)	Positive input from modules B and G
ε(t) = β*δ(t)	Synergistic amplification of module B output by CG_n-P subsystem
if (ε(t) = 0)	Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
ξ(t) = Otx(t)	
else ξ(t) = ε(t)	
if (α = 1)	Repression function inoperative in endoderm but blocks activity elsewhere
η(t) = 0	
else η(t) = ξ(t)	
Θ(t) = γ*η(t)	Final output communicated to BTA

□

1.5.1 Boolean model with logical rules

As mentioned before in the course, the simplest assumption for a gene regulatory network is the boolean approximation, where genes can be either active or inactive (on/off). This can also be interpreted as proteins being present/absent in the cell. Boolean rules (e.g. AND,OR) of the input nodes are defined for determining the state of a node at the next time point. This results in a model with discrete variables and discrete updates in time. The description has the advantage with an enumerable number of possible states for the network, and hence allows for a global exploration of states and dynamics.

Example: boolean description of the lac-operon

In the simple Jacob-Monod model for the lac-operon from the previous example, activity of the operon was determined by presence/absence of lac-repressor and IPTG. In a boolean description the logic of the lac-operon can be described by the following rule

input		output
lac-repressor	IPTG	lac-operon
0	0	1
0	1	1
1	0	0
1	1	1

The only case when the lac-operon is inactive is when the repressor and not the IPTG is present. The repressor is normally expressed. Adding IPTG then causes the lac-operon to switch from inactive to active (as is seen in this model and in previous experiment).□

1.5.2 Michaelis-Menten

The transcription/translation process can be modeled as a transcription factor (TF) binding to DNA (creating a complex) which activates or represses the production of a protein P. A model describing this system is

$$TF + DNA \underset{k_2}{\overset{k_1}{\rightleftharpoons}} TFDNA \overset{k_3}{\rightarrow} P + TFDNA \tag{1.27}$$

Assuming that the binding/release of the transcription factor is fast compared to the production of the protein allows for a Michaelis-Menten formalism to be used. The 'enzyme' in this case is the DNA, and it can be assumed to exist as a single copy within a cell ($DNA + TFDNA = 1$). Solving for the equilibrium of the left part of the reaction leads to $TFDNA = TF/(K + TF)$ where $K = k_2/k_1$. This can be seen as the relative

occupation of the binding site or the fraction of time the transcription factor TF is bound. The production of P can then be seens as this fraction times the rate of production when the regulation is active (given by $k_3 = V_{max}$), which results in

$$\frac{d[P]}{dt} = V_{max}\frac{[TF]}{K + [TF]} \tag{1.28}$$

Note that the reactions described in Eq. 1.27 is not exactly the same as in the Michaeli-Menten enzyme reaction Eq.1.17. How are the parameters V_{max} and K_m defined in this transcription version? When is there no difference compared to the enzymatic case?

Example: Michaelis-Menten repressor

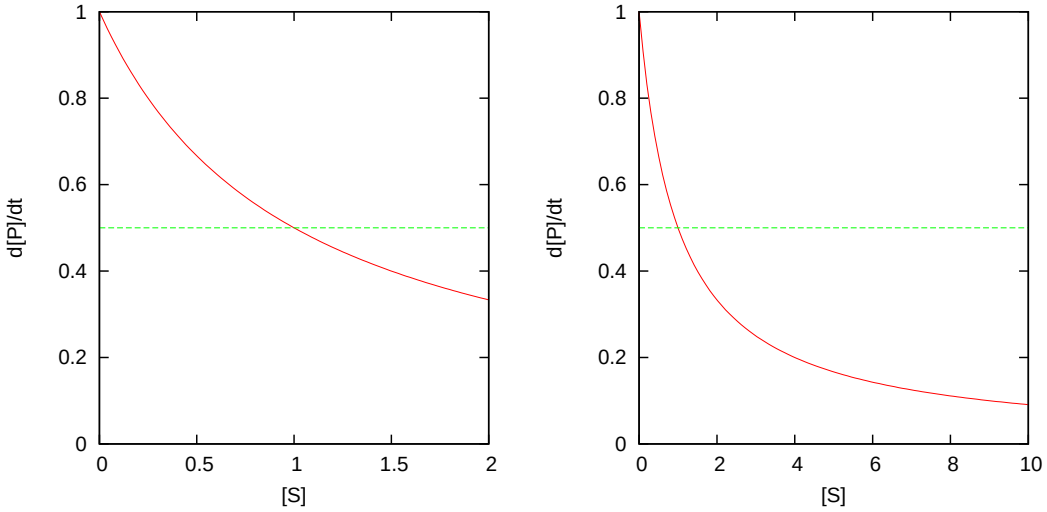
Assume instead that transcription is active if no transcription factor is bound to the DNA, and inactive when the transcription factor (TF) binds

$$\begin{aligned} TF + DNA &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} TFDNA \\ DNA &\overset{k_3}{\rightarrow} P + DNA \end{aligned} \tag{1.29}$$

This leads to a repressor model and working out the Michaelis-Menten formalism (try it!) leads to a production of P described by

$$\frac{d[P]}{dt} = \frac{V_{max}K}{K + [TF]} \tag{1.30}$$

which have the behavior shown in the figure below



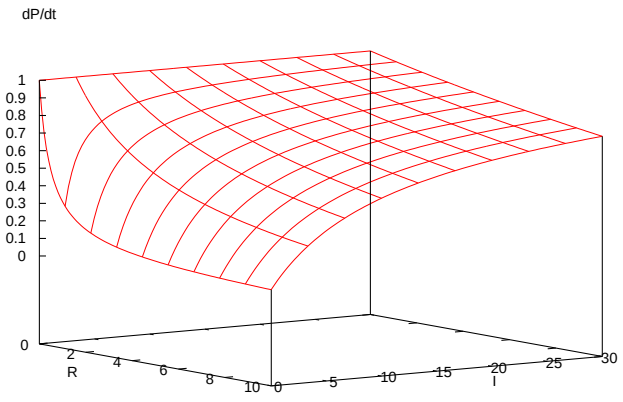
Again this can be seen as the fraction of time the DNA binding site is unoccupied ($K/(K + [TF])$ times the production rate, $k_3 = V_{max}$, when inactive (unoccupied). □

Example: Michaelis-Menten version of the lac-operon

Try to come up with a model of the lac-operon regulation using a Michaelis-Menten formalism for a repressor and mass action kinetics for molecular complex formation of the lac-repressor (R) and IPTG (I). Assume that R binding/unbinding to the DNA is quick in relation to the production, and also that the complex formation between R and I is fast. This leads to an expression of the type

$$\frac{dP}{dt} = \frac{V_{max}K}{K + \frac{K'R'}{K'+I}}, \tag{1.31}$$

where R' is the amount of R free and bound with I. This function is shown in the figure below, and it can be seen that when I is not present R represses the activity, and that the activity increases with increasing concentration of I.



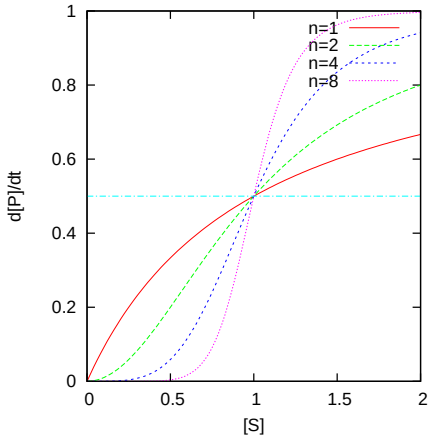
(Note that this example is bordering to be too complicated for this course:) □

1.5.3 Hill-equation

As mentioned in the Michaelis-Menten section on enzyme kinetics, a problem with this formalism is the slow response to changes in substrate concentrations ($\approx$ 81-folded change needed for switching between on/off). For transcription this becomes even more evident, and a common extension of the Michaelis-Menten formalism is the Hill equation. Often it is written in the form

$$\frac{dP}{dt} = V_{max}\frac{S^n}{K^n + S^n} \tag{1.32}$$

where the parameters n and K are called the Hill coefficient and Hill constant, respectively. The Hill constant corresponds to the substrate concentration that results in 50% response, and the Hill coefficient is determining the steepness of the response. The figure below shows the dependance on n given a fixed K.



The Hill-equation can be deduced from a model where a transcription factor can bind to DNA at multiple sites. Hill himself regarded the equation as a model that better fitted experiments, and this is a common standpoint among modelers (i.e. the parameter values are defined by fitting to experiments, rather than from a transcription factor binding model).

Example, Hill from a complex

Assume that two molecules of a single protein type, X, activates the transcription/translation of another protein, P. The reactions can be formulated as

$$X + X + DNA \underset{k_2}{\overset{k_1}{\rightleftharpoons}} TFDNA \overset{k_3}{\rightarrow} P + TFDNA \tag{1.33}$$

From the equilibrium of the left reaction (together with the assumption $DNA+TFDNA = 1$), the fractional occupancy of the binding site is given by $TFDNA = X^2/(K + X^2)$, where $K = k_2/k_1$ (show this!). The production rate is then determined by ($k_3 = V_{max}$)

$$\frac{dP}{dt} = V_{max}\frac{X^2}{K + X^2}. \tag{1.34}$$

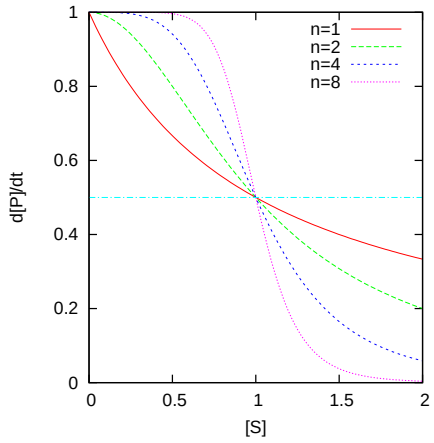
□

Example, Hill repressor

In the case of a repressor S deactivating the transcription of P, the Hill-equation looks like

$$\frac{dP}{dt} = V_{max}\frac{K}{K+S^n} \tag{1.35}$$

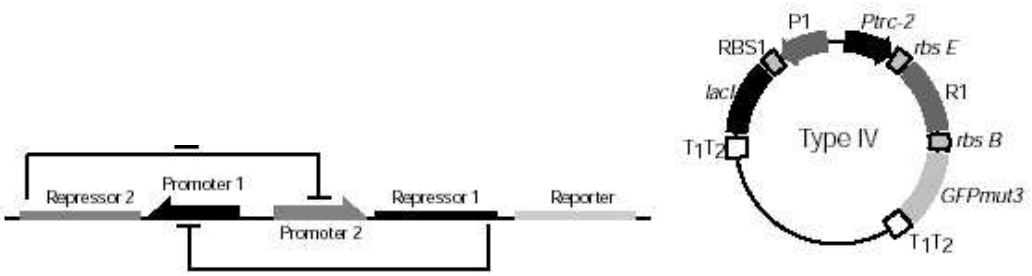
which shows a n dependance as in the figure below.



□

Example, bistable switch

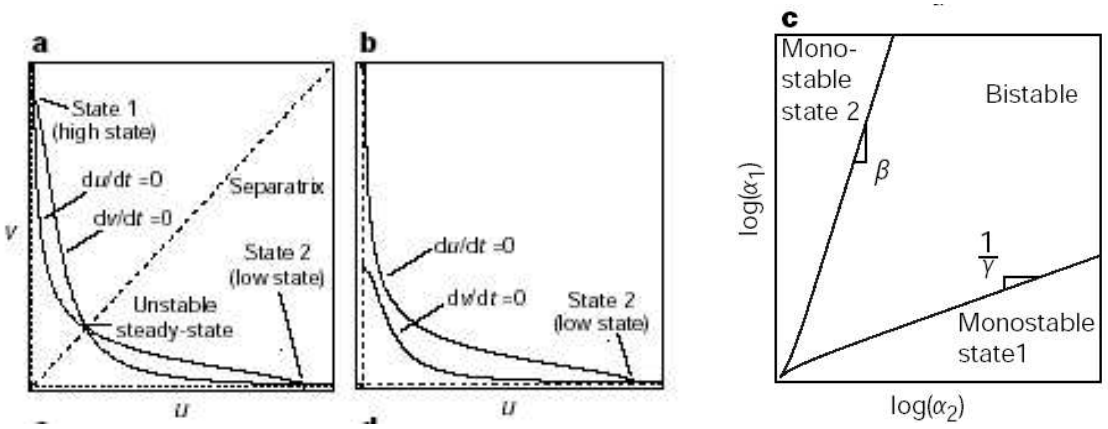
In a beautiful work by Gardner et.al. a genetic switch is created by direct manipulation of the DNA in *E. coli* (figure below). A network of two genes repressing each other is constructed, and this novel technique allows for creating simple systems where direct comparisons between models and experiments are more tractable.



The equations used in this model are of Hill-type plus addition of a constant degradation term.

$$\begin{aligned} \frac{du}{dt} &= \frac{\alpha_1}{1+v^\beta} - u \\ \frac{dv}{dt} &= \frac{\alpha_2}{1+u^\gamma} - v \end{aligned} \tag{1.36}$$

The model can behave as a bistable switch where two stable fixed points are defined by (u, v)=(high,low) and (low,high) respectively. A phase plane plot with the nullclines (calculate them!) are shown in the figure below, and quite interestingly, either β or γ needs to be larger than one to get the bistable behavior. Otherwise the system has a single stable fixed point.



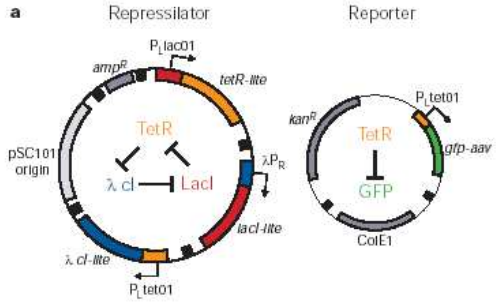
□

1.5.4 Models accounting for both transcription and translation

Sofar, we have only looked at models describing the transcription and translation in a single equation. It is of course also possible to divide these into two different processes, and also treat the mRNA as a dynamical variable.

The repressilator

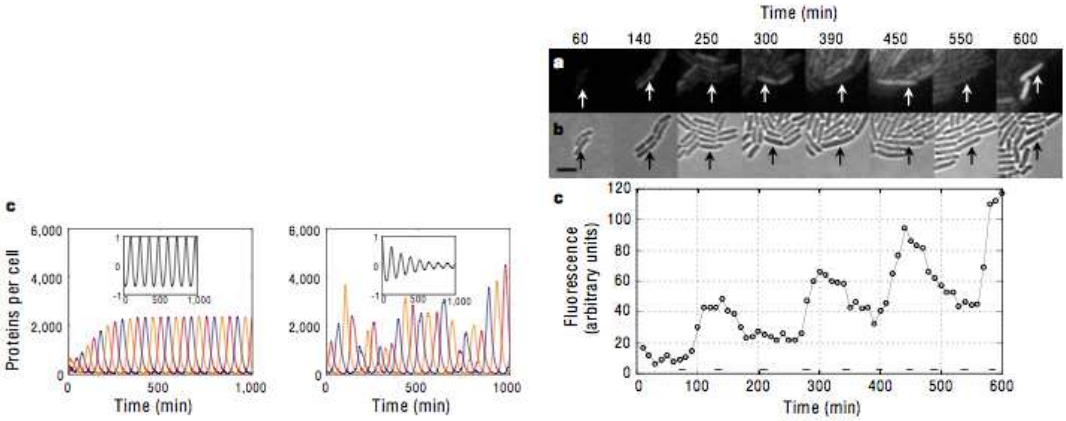
In a similar effort as described in the bistable switch example, Elowitz et.al. constructed a network of three repressing genes (figure). A computer exercise is devoted to modeling of this system, and details are left for then, but the equations used are presented below as an example of a transcription/translation model.



$$\frac{dm_i}{dt} = -m_i + \frac{\alpha}{(1+p_j^n)} + \alpha_0 \qquad \begin{pmatrix} i = lacI, tetR, cI \\ j = cI, lacI, tetR \end{pmatrix}$$

$$\frac{dp_i}{dt} = -\beta(p_i - m_i)$$

The m variables represent mRNA and the p variabless represent proteins. The transcription is modeled by a Hill-type equation, and translation is modeled by a linear equation. In addition to this, constant degradation of all molecules are modeled. The figure below show the oscillating behavior achieved both in the simulations, and in the experiments. The left simulation plot shows the deterministic model described above, and the right plot shows a stochastic version.



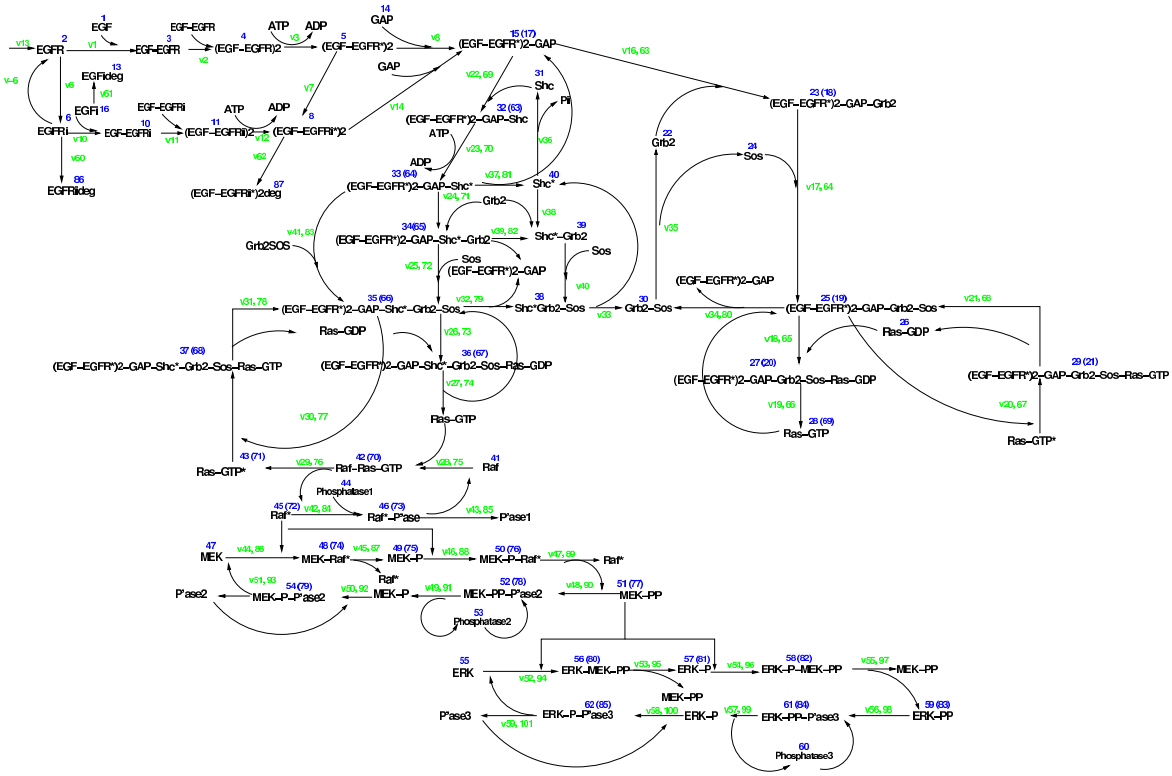
□

1.6 Large molecular networks; systems biology in a nutshell

Using the building blocks of mass action and enzymatic reactions, and transcription/translation descriptions, models of large biochemical networks can be developed. In these cases analytical solutions are unreachable, and computer simulations of the systems are necessary.

Example: EGF-pathway simulation

The receptor to the epidermal growth factor (EGF) ligand belongs to the tyrosine kinase family of receptors and is expressed in virtually all organs of mammals. EGF receptors play a complex role during development and in the progression of tumors. Schoeberl *et.al.* have created a model of the pathway as shown in the figure below.



This might look like a far too advanced example for our purposes, but let's look at the reaction for a single molecule, e.g. the Raf. It is directly involved in two reactions

$$\begin{aligned} Raf + RasGTP &\underset{k_{-28}}{\overset{k_{28}}{\rightleftharpoons}} RafRasGTP \\ Raf^*P1 &\overset{k_{43}}{\rightarrow} Raf + P1 \end{aligned} \tag{1.37}$$

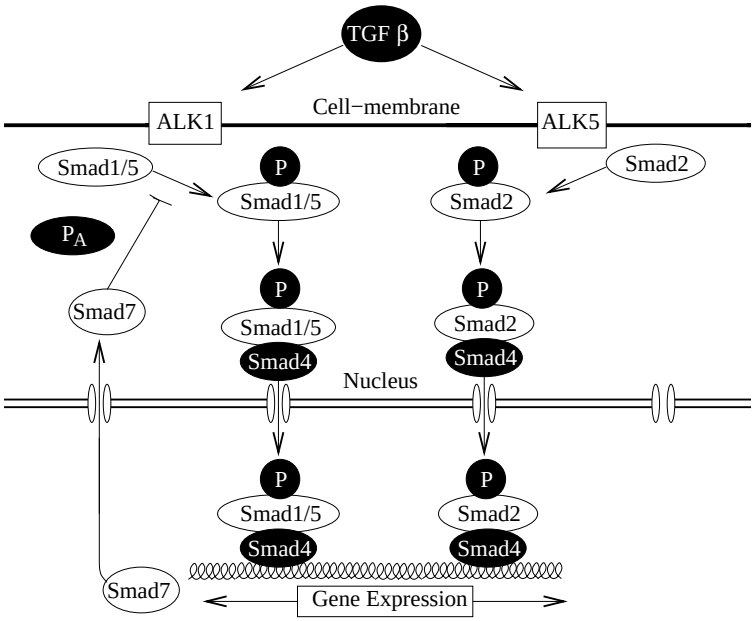
and the formulation of the differential equation for Raf is straightforward using the mass action formalism

$$\frac{d[Raf]}{dt} = -k_{28}[Raf][RasGTP] + k_{-28}[RafRasGTP] + k_{43}[Raf^*P1] \tag{1.38}$$

□

Example: TGF-β pathway

The TGF-β pathway plays a prominent role in inter- and intracellular communication and subversion can lead to cancer, fibrosis vascular disorders and immune diseases.



This network includes both molecular reactions and transcriptional regulation. A model for the pathway can be defined by the following reactions:

$$\emptyset \underset{p_0p_1}{\overset{p_0}{\rightleftharpoons}} \text{ALK1} \quad (1)$$

$$\emptyset \underset{p_4p_5}{\overset{p_4}{\rightleftharpoons}} \text{Smad4} \quad (2)$$

$$\emptyset \underset{p_8p_9}{\overset{p_8}{\rightleftharpoons}} \text{ALK5} \quad (3)$$

$$\text{TGF}\beta + \text{ALK1} \underset{p_{14}}{\overset{p_{13}}{\rightleftharpoons}} \text{TA1} \quad (4)$$

$$\text{PSmad1} + \text{Smad4} \underset{p_{19}}{\overset{p_{18}}{\rightleftharpoons}} \text{PS14} \quad (5)$$

$$\text{TGF}\beta + \text{ALK5} \underset{p_{21}}{\overset{p_{20}}{\rightleftharpoons}} \text{TA5} \quad (6)$$

$$\text{P}_\text{A} + \text{TA1} \underset{p_{28}}{\overset{\text{Smad7}\ p_{27}}{\rightleftharpoons}} \text{TA1P} \quad (7)$$

$$\text{P}_\text{B} + \text{TA5} \underset{p_{32}}{\overset{\text{Smad7}\ p_{31}}{\rightleftharpoons}} \text{TA2P} \quad (8)$$

$$\emptyset \underset{p_2p_3}{\overset{p_2}{\rightleftharpoons}} \text{Smad1} \quad (9)$$

$$\emptyset \underset{p_6p_7}{\overset{p_6}{\rightleftharpoons}} \text{Smad2} \quad (10)$$

$$\emptyset \underset{p_{10}}{\overset{PS14N\ (p_{11},p_{12})}{\rightleftharpoons}} \text{Smad7} \quad (11)$$

$$\text{Smad1} \underset{p_{17}}{\overset{TA1\ (p_{15},p_{16})}{\rightleftharpoons}} \text{PSmad1} \quad (12)$$

$$\text{Smad2} \underset{p_{24}}{\overset{TA5\ (p_{22},p_{23})}{\rightleftharpoons}} \text{PSmad2} \quad (13)$$

$$\text{PSmad2} + \text{Smad4} \underset{p_{26}}{\overset{p_{25}}{\rightleftharpoons}} \text{PS24} \quad (14)$$

$$\text{PS14} \underset{k_{30}}{\overset{p_{29}}{\rightleftharpoons}} \text{PS14N} \quad (15)$$

Table 1.1: The different reactions in the TGF-β pathway model, where p_i $(i = 0, 1, \ldots, 32)$ are the rate constants. Reactions with the symbol $\emptyset$ model production and degradation. In reactions (11), (12) and (13) Michaelis-Menten dynamics is used.

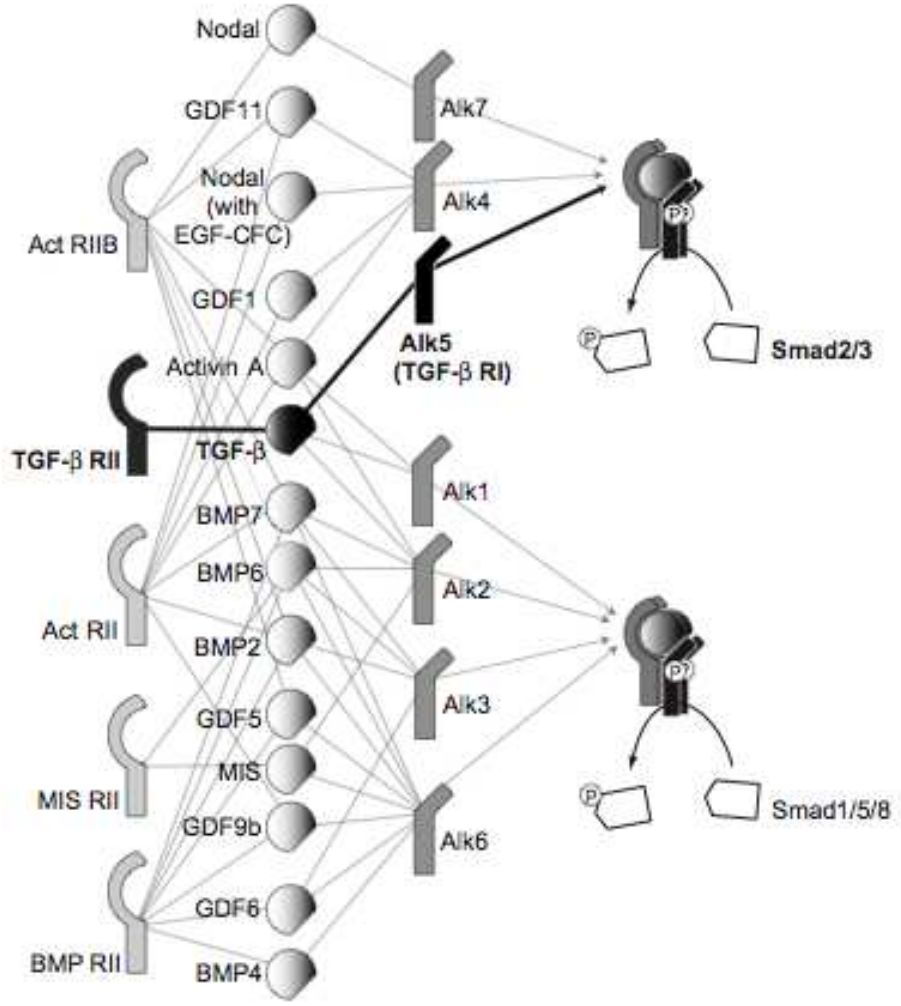
As an example, the model equation for the Smad1 concentration is given by

$$\frac{d[\text{Smad1}]}{dt} = p_2 - p_2p_3[\text{Smad1}] + p_{17}[\text{PSmad1}] - \frac{p_{15}[\text{Smad1}][\text{TA1}]}{p_{16} + [\text{Smad1}]}, \quad (1.40)$$

which is extracted from reactions 9 and 12 above. Try to extract the model equation for another molecule! □

Example: The TGF-β family of ligands and their receptors

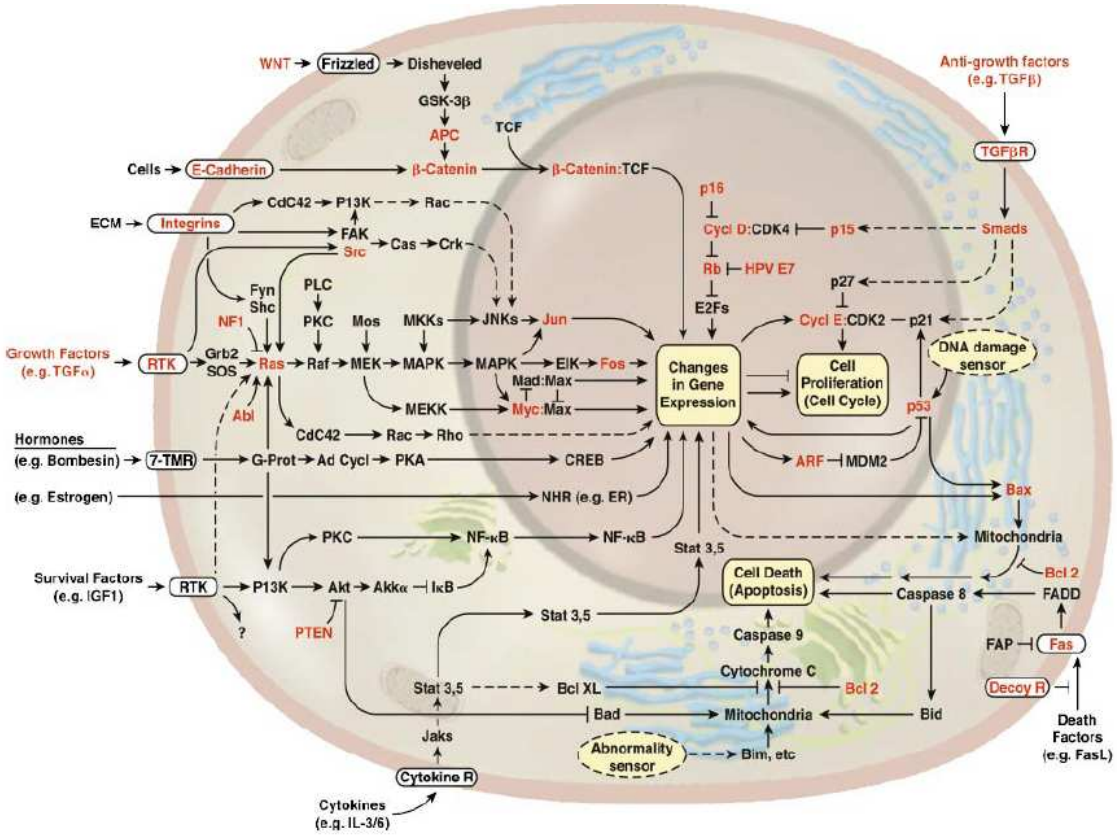
In the previous example, a module of the TGF-β signalling pathway was presented. In idealized experiments, this module can be investigated. A problem that might have to be accounted for in a modeling approach is crosstalk between a model and its surrounding (all molecules left out of the model). Hence the presented model might not correctly describe the behavior within a living organism. For example, TGF-β is only one member of a whole family of ligands, that binds to a number of different receptors and each ligand-receptor combination can activate/deactivate the same pathway (see figure).



□

Example: Pathways of relevance for cancer

The complexity within living cells are even larger than shown in the previous examples. Both the EGF and the TGF-β pathways are important at cancer progression. As shown in the figure below (from Carstens introduction), these pathways are only two of multiple pathways that are important in this case.



This is an example of a number of modules (the specific pathways with robust behavior and 'output') that interact with each other.□

1.7 Estimation of parameter values

Even when the mathematical description of a model is defined (as in the previous section) the dynamical behavior can change due to different values of the parameters. A main task withinin a modeling approach is to find or estimate parameter values that are relevant for the biological system at hand. Here we will discuss two diffrent approaches for estimating parameter values; experimental measurements, and reverse engineering.

1.7.1 Experimentally measuring parameter values

If it is possible, a good way to find parameter values is to measure the dynamics of a single reaction. From this it is then possible to estimate the rate parameters.

Example: ALK1 internalization rate

In an experiment, the ALK1 receptor at the cell membrane is labeled with an antibody, and after 15 minutes the amount of labeled ALK1 receptor is measured. At this time only 5% of the labeled ALK1 molecules are still present.

Assume a reaction $X \xrightarrow{k} \emptyset$ as the receptor disappears from the membrane, which leads to an equation

$$\frac{dX}{dt} = -kX. \tag{1.41}$$

The solution to this equation is $X(t) = X_0 e^{-kt}$ where X_0 is the initial concentration. (This is easily checked by taking the time derivative of $X(t)$.) The kinetic parameter can be estimated by

$$\begin{aligned} e^{-kt} &= \frac{X(t)}{X_0} = 0.05 \\ k &= -\frac{1}{t}\ln\frac{X(t)}{X_0} = -\frac{1}{15}\ln 0.05 = 0.2 \text{ min}^{-1} \end{aligned} \tag{1.42}$$

This estimate could be improved further by fitting a curve $X = X_0 e^{-kt}$ to a dynamical measurment of the labeled ALK1 receptors.□

1.7.2 Reverse engineering

Even if parameter values are not known from experiment it can be possible to do a reverse engineering to find parameters for the model that result in an agreement of model and some biological features of the system. The first thing needed is an objective function (error measure) that is a quantitative measure of how well the model behavior (for a given parameter set) corresponds to the biological feature at hand. Then an optimization method is needed to find parameters that result in an optimal value of the objective function. Typically this a hard optimization problem in a high dimensional parameter space, and one has to rely on iterative heuristic algorithms to find 'good' solutions.

Objective function

The objective function, $R(p)$, is a function of the model parameters p. If the system of differential equations for the model is not analytically solvable, a simulation of the model for specific parameters is needed for evaluating the objective function value. The most

common type of objective function assumes that there are some quantitative experimental data available for molecular concentrations allowing for a direct comparison with the model variables. If for example the concentration of protein X has been measured at N time points $t_1, t_2, ..., t_N$, a mean square error can be defined as

$$R(p) = \frac{1}{N}\sum_{t}^{N}(X_t^{\text{exp}} - X(p)_t^{\text{model}})^2 \tag{1.43}$$

where X_t^{exp} are the measured concentrations and $X(p)_t^{\text{model}}$ are the model variable values at different time points.

Optimization algorithms

Typically, iterative algorithms are used when optimizing the objective function. The iterative procedure consists of three steps: 1) Solve the differential equation and calculate the objective function value, 2) Adjust model parameters (from a random selection) and resimulate, 3) Accept or reject the new parameters depending on the difference in objective function value.

Three examples of optimization algorithms that could be used for parameter estimations are

- **Local search.** This is the naive way of trying to find a good value for the objective function. Here you start with a parameter set for which the model is simulated and the objective function is evaluated. After adjusting parameters a new objective function value is evaluated and it is accepted if this value is lower than the previous one. This means that we will only go downhill in the objective function 'landscape' and we will end up in the closest local minimum.

- **Simulated annealing.** Again, you start with a parameter set for which the objective function is evaluated, then do a parameter adjustment and reevaluate the objective function. Now the new parameter set is accepted with a probability one if $\Delta R = R_{new} - R_{old}$ is negative, and with probability $e^{-\Delta R/T}$ if ΔR is positive. T is a parameter (fictitious temperature) which tunes the probability. The first thing to note is that the algorithm can allow for accepting new parameter sets with a higher objective function value, which means that it can escape from local minima. The second thing to note is that at high values of T, almost all parameter adjustments are accepted and we get something like a random walk in the parameter space (searching large regions). At low values of T almost only decreased objective function values are accepted. The algorithm starts at high values of T and then slowly decreases T until no more updates are accepted.

- **Genetic algorithms.** This type of algorithm is developed from an evolutionary fitness principle. It starts with an ensemble of parameter values for which the objective function is evaluated. Then 'good' parameter sets are kept, 'bad' ones are

removed. The bad solutions are replaced by forming new parameter sets from two principles; mutation, where the parameters of a good solution are slightly adjusted, and mating, where the new parameter set is some kind of combination of two good solutions.

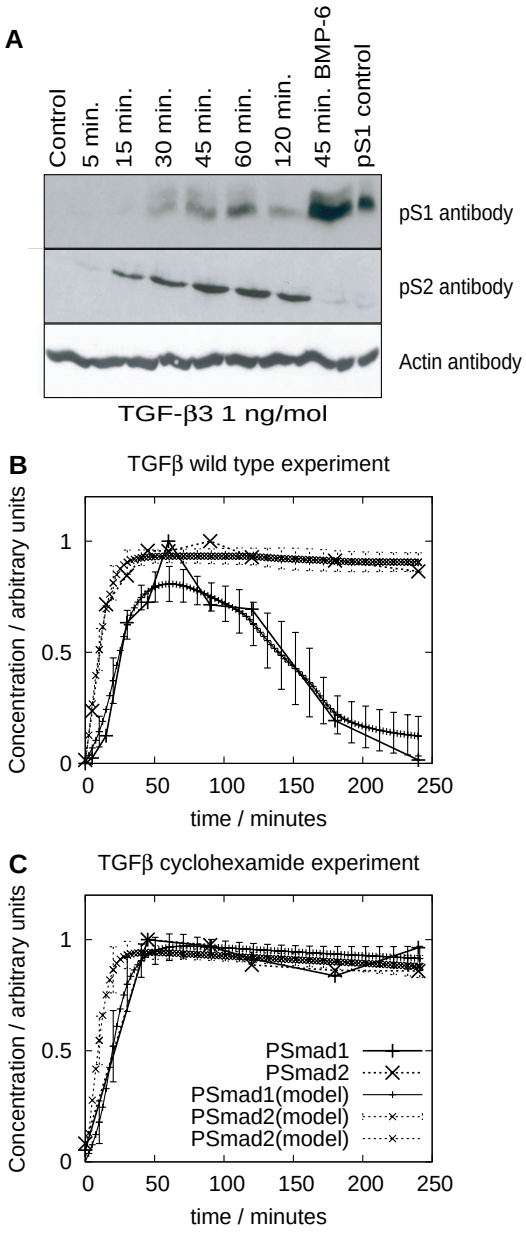
Example: TGF-β model

For the TGF-β pathway, PSmad1 and PSmad2 concentrations are measured at different times after TGF-β stimulation. The concentrations are measured at N discrete time points $t_1, t_2, \ldots t_N$ for two experiments. The model is optimized using simulated annealing type of algorithm and the mean square error is used as an objective function:

$$R(\mathbf{p}) = \frac{1}{N}\frac{1}{M}\sum_{t=t_1}^{t_N}\sum_{i=1}^{M}(x_i(t) - \tilde{x}_i(t))^2, \tag{1.44}$$

where $x_i(t, \mathbf{p})$ and $\tilde{x}_i(t)$ denote model points and experimental points respectively and the index i denotes the different molecules ($M = 2$ in total). (The sum of the R values from the two experiments is used as objective function.)

The figure below shows experimental data, and the model output for optimized parameters. In this case multiple good solutions were found (the average model behavior is plotted with errorbars).



□

1.8 Model analysis in systems biology

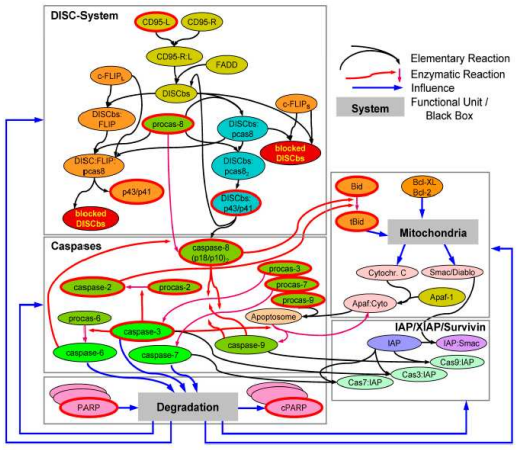
1.8.1 Robustness

Biological systems are often very robust to fluctuations in the environment. Also, the same modules (e.g. pathways) exist in many different species with different environment.

A good model should also reflect this and hence a test for robustness can be an important test of the model. Robustness analysis can also pinpoint which reactions/parameters that are important for obtaining a specific biological behavior. In section 1.3.3 the concept of sensitivity analysis was introduced, and here some examples are given where it has been applied for biological networks.

Example: CD95-induced apoptosis

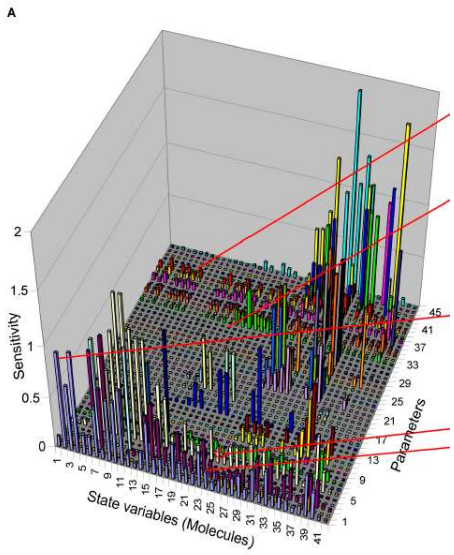
This model developed by Bentele et.al. describes a pathway that regulates apoptosis (programmed cell death). Defects in the regulation of apoptosis result in serious deseases such as cancer, autoimmunity and neurodegeneration. The model components are shown in the figure below.



A local sensitivity analysis is applied to a single solution (parameter set). The measure used is the integral of the protein concentration $c_i = \int_t x_i dt$ where x_i is a concentration. In the figure below the absolute value of the sensitivities,

$$s_{ij} = \frac{dc_i/c_i}{dp_j/p_j}, \tag{1.45}$$

are shown for all molecules i and parameters j.



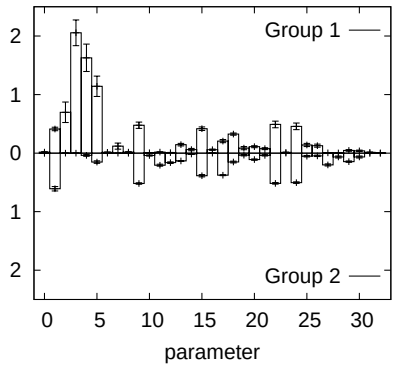
□

A problem with the local sensitivity measure is that it can be very dependent on the parameter values. One way to improve the sensitivity measure is to measure the local sensitivity in multiple points spanning a region in the parameter space.

Example: TGF-β model

For the TGF-β model the optimization provided multiple solutions that could explain the experimental data (as shown in a previous example). These solutions can be grouped into those that utilizes the Smad7 feedback and those that do not.

The figure below shows average sensitivity measures calculated from multiple solutions for each group. The sensitivity is measured on the integral of PSmad1 and PSmad2 concentrations for each parameter. The soutions in group 2 (those using Smad7 feedback) are more robust.



□

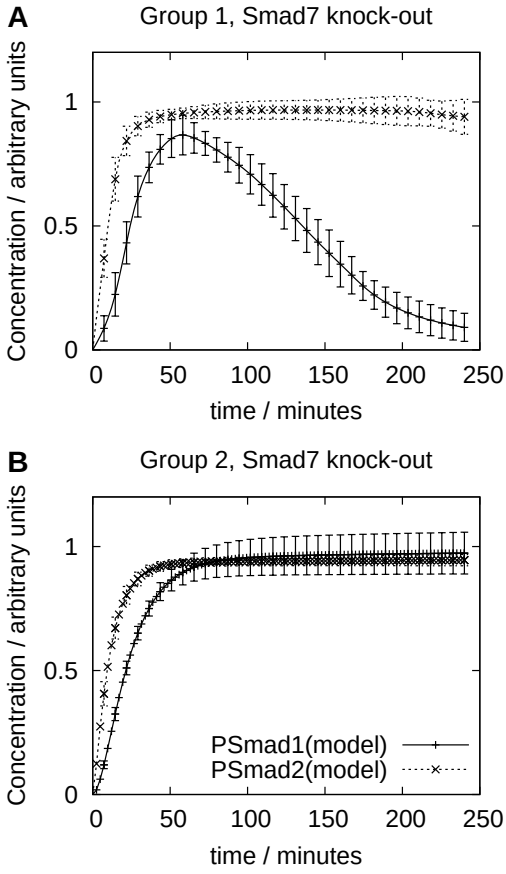
There are other means to measure more global robustness, which will be discussed in the last lecture.

1.8.2 Perturbations

Another way of analysing a model is by doing perturbations, such as e.g. removing a molecule in the model. The model behavior could then be compared to the same perturbation in experiments, or predict new biology. The main benefit of having a model in this case is that perturbations are easy to do in the model, while it is often long and hard work to do it experimentally. Multiple perturbations can be tested in a model framework, and those that results in interesting behavior could then be tested in experiments.

Example: Perturbation in the TGF-β model

The figure below shows the model predictions if Smad7 is removed from the TGF-β model. Again it is shown for two groups of solutions, where the two groups provide different predictions.



□

1.9 Transport

Reactions within a cell occur at different spatial locations. For example, a signal transduction network usually have reactions at the cell membrane, in the cytoplasm, and in the nucleus. Hence spatial dynamics of molecules might also be important for the behavior of a biochemical network within a cell. Spatial considerations become even more important when modeling multicellular systems, where it is known that signalling molecules (often termed morphogens) can be produced at specific positions, move out in the surrounding tissue, and regulate development.

1.9.1 Diffusion

Molecules are constantly moving and bouncing into each other due to thermal effects. This Brownian motion leads to molecular diffusion. Consider a microscopical model for diffusion that describes number of molecules on a one dimensional lattice discretized in

time (x_i, t_k), where $x_{i+1} - x_i = \Delta x$ and $t_{k+1} - t_k = \Delta t$. The number of molecules in position x_i at time t_k is denoted n_i^k. Assume that each molecule moves Δx either to the right or to the left during a time Δt with probabilities $P_l = P_r = 1/2$. Also assume that consecutive moves are uncorrelated. The average change in molecular number at a spatial point x_i in a time step Δt is given by

$$\begin{aligned} n_i^{k+1} - n_i^k &= \Delta n_i^k = P_r n_{i-1}^k - (P_l + P_r) n_i^k + P_l n_{i+1}^k \\ &= \frac{1}{2} n_{i-1}^k - n_i^k + \frac{1}{2} n_{i+1}^k = \frac{1}{2} \left(n_{i-1}^k - 2n_i^k + n_{i+1}^k \right) \\ &= \frac{\Delta x^2}{2} \frac{n_{i-1}^k - 2n_i^k + n_{i+1}^k}{\Delta x^2} \end{aligned} \tag{1.46}$$

This leads to a change per Δt as

$$\frac{\Delta n_i^k}{\Delta t} = \frac{\Delta x^2}{2\Delta t} \frac{n_{i-1}^k - 2n_i^k + n_{i+1}^k}{\Delta x^2} = D \frac{n_{i-1}^k - 2n_i^k + n_{i+1}^k}{\Delta x^2} \tag{1.47}$$

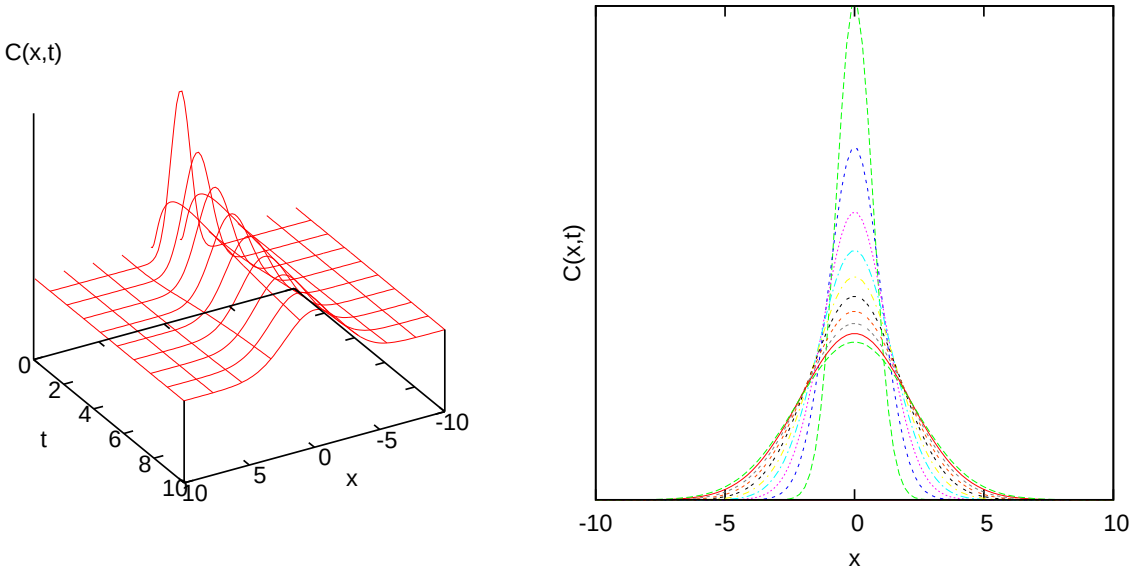
where $D = (\Delta x)^2/(2\Delta t)$ is defined as the diffusion constant. The experienced reader can recognize that the right hand side of the equation corresponds to a discrete version of the second derivative in x ($\approx d^2n/dx^2$). Letting $\Delta x \to 0$ and $\Delta t \to 0$ while keeping D constant, and transforming number of molecules into concentrations, C ($C = n/vol$) leads to

$$\frac{dC}{dt} = D \frac{d^2C}{dx^2} \tag{1.48}$$

which is Fick's law. This is a partial differential equation in time and space and describes diffusion in a continuous setting. Solving it is beyond the scope of this course.

Example: diffusion from a peaked distribution

A concentration peaked at a single point in space will diffuse as shown in the figure



□

Example: diffusion times

The time it takes for a diffusive substrate to "reach" a distance L can be approximated by

$$t = \frac{L^2}{2D} \tag{1.49}$$

The value of the diffusion constant, D, for a small molecule (e.g. glucose) is in the order of 10^{-9} m^2/s. Given a cell size of $L \approx 50$ μm the diffusion time within a cell is approximately

$$t = \frac{(50 \times 10^{-6})^2}{2 \times 10^{-9}} \approx 3 \text{ s} \tag{1.50}$$

while a macroscopic length as $L = 1$m would give

$$t = \frac{(1)^2}{2 \times 10^{-9}} \approx 5 \times 10^8 s \approx 16 \text{ years !} \tag{1.51}$$

□

If diffusion is included in a model, it can be integrated as an ordinary differential equation on a discretized space, where the formulation is

$$\frac{dC_i}{dt} = D \left(\sum_j^{N_{neigh}} (C_j - C_i) \right) \tag{1.52}$$

where i is a compartment index and the sum over j is the N_{neigh} neighbors. Here the distances and cross section areas between compartments are assumed to be equal and included in the diffusion constant D.

Example: diffusion between two compartments

The diffusion rate is proportional to the molecular concentration, similar to what is given for a mass action reaction. This is particular apparent in the case of two copartments where the spatial factors are incorporated in the diffusion constant. Assume diffusion of molecule A between compartments i and j.

$$A_i \underset{D}{\overset{D}{\rightleftharpoons}} A_j \tag{1.53}$$

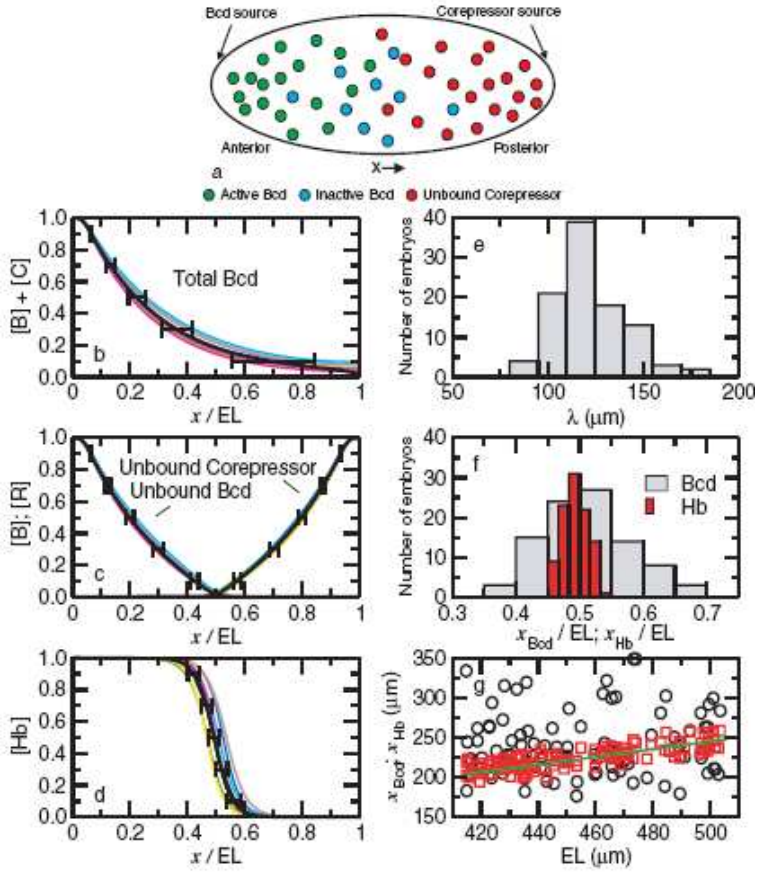
The resulting differential equations are given by

$$\frac{dA_i}{dt} = -\frac{dA_j}{dt} = -DA_i + DA_j \tag{1.54}$$

□

Example: early patterning in *Drosophila*

Diffusing signalling molecules (morphogens) are important for regulating development in multicellular organisms. In the *Drosophila* embryo, bicoid mRNA is deposited at the anterior pole (a localized source). This model by Howard *et.al.* (2005) discuss how this robustly can lead to a very precise gene expression pattern.



For the interested reader, the model equations are provided.

$$\frac{\partial[B]}{\partial t} = D\frac{\partial^2[B]}{\partial x^2} - \mu[B] - \nu[B][R] + J_B\delta(x - x_B(t))$$

$$\frac{\partial[R]}{\partial t} = D\frac{\partial^2[R]}{\partial x^2} - \mu[R] - \nu[B][R] + J_R\delta(x - x_R(t))$$

$$\frac{\partial[C]}{\partial t} = D\frac{\partial^2[C]}{\partial x^2} - \mu[C] + \nu[B][R]$$

$$\frac{\partial[hb]}{\partial t} = D_{hb}\frac{\partial^2[hb]}{\partial x^2} - \mu_{hb}[hb] + \frac{\beta[B]^3(\eta[\mathrm{Hb}] + \gamma K)}{K^4 + [B]^3(\eta[\mathrm{Hb}] + K)}$$

$$\frac{\partial[\mathrm{Hb}]}{\partial t} = D_{\mathrm{Hb}}\frac{\partial^2[\mathrm{Hb}]}{\partial x^2} - \mu_{\mathrm{Hb}}[\mathrm{Hb}] + \alpha[hb].$$

□

1.9.2 Reaction-Diffusion models

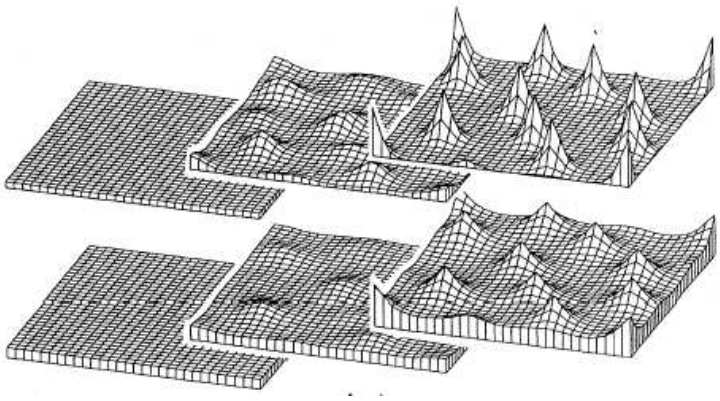
Models combining biochemical reactions and diffusion have the ability to create spatial patterns in molecular concentrations. This was first noted by Turing in the 1950s.

Example: the activator-inhibitor model

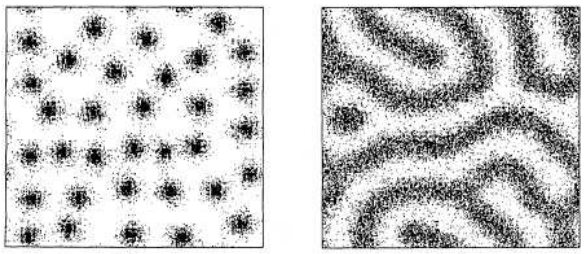
Meinhardt introduced an activator (a) inhibitor (h) reaction-diffusion model. The one dimensional version of the equations look like

$$\begin{aligned} \frac{da}{dt} &= \rho_a \left(\frac{a^2}{h} - a \right) + D_a \nabla^2 a \\ \frac{dh}{dt} &= \rho_b (a^2 - h) + D_h \nabla^2 h \end{aligned} \tag{1.55}$$

The figure shows the spontanous pattern formation in activator (top) and inhibitor (bottom) concentrations when starting in a close to homogeneous state.



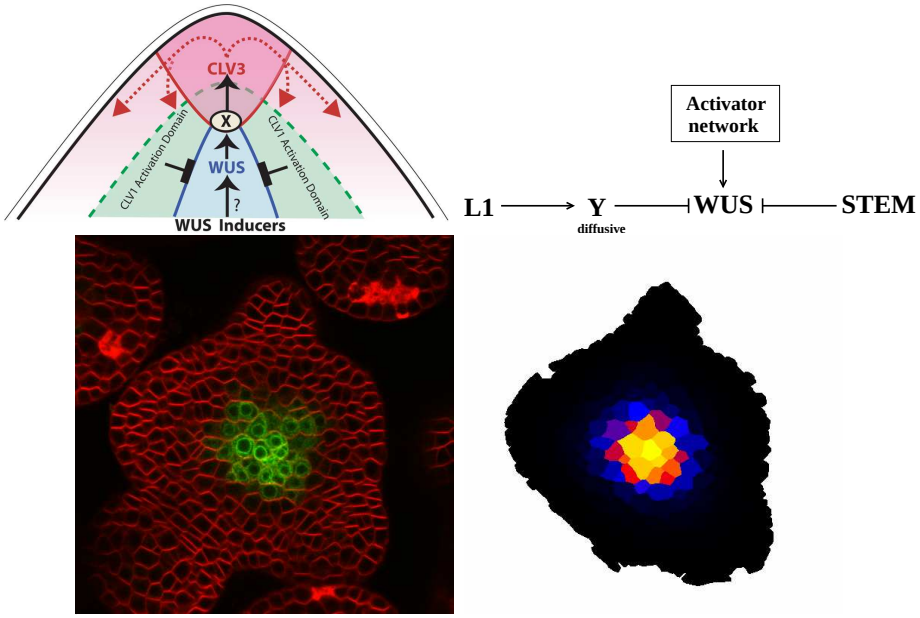
Different types of patterns of the activator, generated from different parameter sets, are shown in the figure below.



□

Example: stem cell regulation in plants

At the tip of a plant shoot, there is a pool of stem cells throughout the adult life of the plant. These cells are in part regulated by the WUS protein which is expressed in the interior of the shoot (see figure). This expression is very robust, and even removal of the shoot will lead to a new WUS domain forming. A model in which WUS is assumed to be induced by an activator network is capable of explaining this ability of reorganization.



For completeness, the equations are provided.

$$\frac{dW}{dt} = \frac{1}{\tau_w} g\left(h_w + T_{wa}A + T_{wy}Y\right) - d_w W \tag{1.56}$$

$$\frac{dY}{dt} = k_y L_1 - d_y Y + D_y \nabla^2 Y \tag{1.57}$$

$$\frac{dA}{dt} = a - (b+\beta)A + cA^2B - dYA + D_a \nabla^2 A \tag{1.58}$$

$$\frac{dB}{dt} = bA - cA^2B + D_b \nabla^2 B. \tag{1.59}$$

where $g(x)$ is the sigmoidal function

$$g(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{1+x^2}}\right). \tag{1.60}$$

The parameter τ_i is the inverse maximal rate, and h_i sets the basal expression level. The T_{ij} parameters define the strength of the regulation (j regulating i). A positive T defines an activation, while a negative T leads to a repression.□