Mathematical models in systems biology

Models in biology

The term "model" has many uses in biology. We speak of model organisms, structural models, cellular models etc. But, what exactly are models and why do we use them?

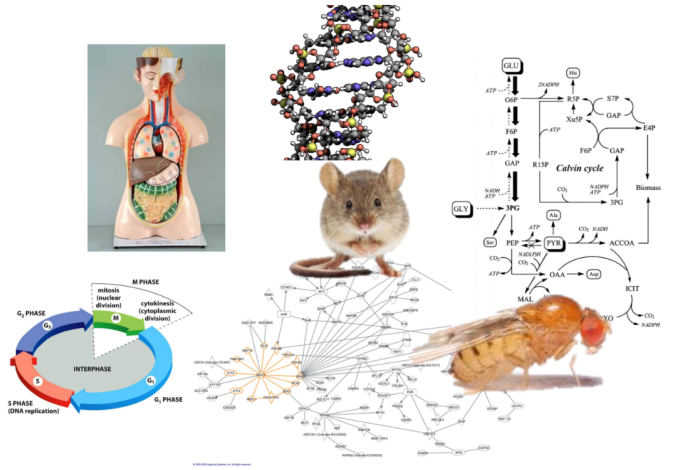
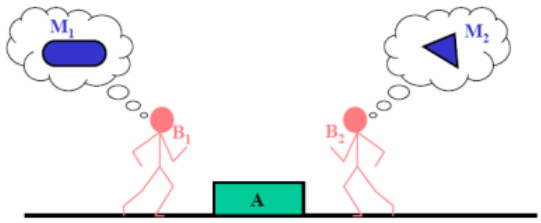


Figure 1 *Examples of models used in biology.*

The following, rather abstract, definition of a model was given by Marvin Minsky (a computer scientist and co-founder of the MIT artificial intelligence lab):

"To an observer B an object M is a model of an object A to the extent that B can use M to answer questions that interest him about A."



One interesting aspect of this definition is that the model of an object is not only shaped by the object itself. To a very large extent, it is shaped by the questions asked about the object. So models can differ even though they describe the same object and we may choose different models depending on our questions and on the information we have about the object. We will see examples of this throughout the course.

While there do exist some objective criteria for evaluating the appropriateness of a certain model, the type of model one chooses also depends on personal preferences. It is therefore even possible for two individuals to ask the same question about an object but still use different models of the object.

One aspect of models emphasized by Minsky's definition is the idea of utility. A model should be useful for answering questions about an object. Thus, a model should be built with a specific purpose or question in mind. The quality of a model should then be assessed primarily by its usefulness for answering the questions at issue. This view was expressed concisely by George P. Box: "All models are wrong, but some are useful."

Types of models in biology

Grouping models into different model types allows us to discuss the advantages and disadvantages of different kinds of models. In doing so, we should not forget that these different kinds of models actually form a continuum. The types proposed are not always clear cut and the grouping inevitably somewhat arbitrary.

The least formal type of models we use in biology is the **verbal model.** In this type of model, a process, system, or object is described by natural language (e.g., plain English). Such verbal models are often the first type of model we develop in order to describe a system. "Kinase X phosphorylates protein Y which in turn activates the expression of gene Z" is a simple example of such a model. The analysis of verbal models is based largely on common sense and intuition. The relative simplicity of verbal models and the lack of formal methods for their analysis do not necessarily mean that such models are less useful than more formalized models. In fact, there are many historical examples of verbal models that have enabled major breakthroughs in biology. Notable examples are Mendel's "laws" of inheritance, Robert Koch's microbial model of disease and the "central dogma" of molecular biology (DNA -> RNA -> protein).

Natural language usually leaves lots of room for ambiguity and of course verbal models are no exception. So as we try to make models more rigorous and precise, natural language becomes increasingly cumbersome. There comes a point where using **semi-formal or formal information models makes sense** (Figure 2)**.** These models replace natural language with well-defined (i.e., formalized) language and thereby reduce ambiguity. Instead of words, formalized languages often use graphical symbols to represent the relationship between different components of a model.

Mathematical models are the most formalized type of models. The relationship between different components of a model and the state of these components is described through mathematical equations. The key advantage of these mathematical models is that once relationships are

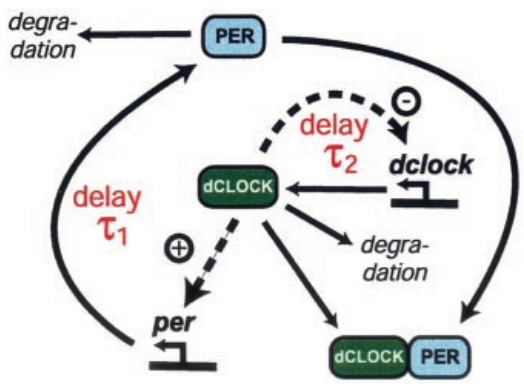


Figure 2 *Example of a semi-formal information model in which the relationship between the different components is represented by different types of arrows. As is often the case in semi-formal models, the arrows indicate information flow. Note that the arrows indicate the direction of the flow but say nothing about type or quantity of information. (adapted from Smolen et al. Biophys J. (2002))*

cast in a mathematical form, the full range of mathematical techniques can be used to analyze them. However, to be able to describe relationships in mathematical terms in the first place, we need detailed and precise information about the system we are trying to model.

This systems biology concept course will deal mostly with mathematical models as well as semi-formal and formal information models.

Another useful way to classify models is by the complexity of the networks that are being modeled (i.e., the number of model components) and by the level of information (detail/accuracy) that is required/available for each component and inter-component relationship (Figure 3).

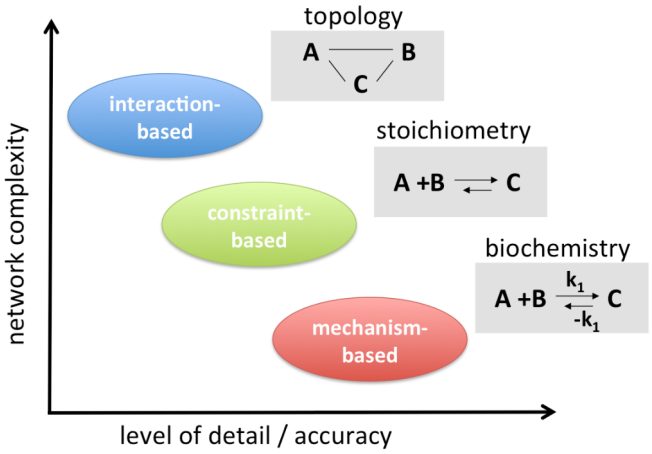


Figure 3 *Classification of the three types of models you will encounter in this course (colored ovals). The position on the plot indicates the relative complexity of the model and the level of detail/accuracy of these models. The grey boxes give examples.*

In **interaction-based models,** the description of the relationship between model components tends to be highly simplified. For example, the relationship is represented by a simple binary value. Either two components interact, or they do not.

This simplicity permits the analysis of very large models with many network components. The amount of data required to describe each relationship is very small. You will have a closer look at such interaction-based models in the second part of the course during Challenge 2.

Mechanism-based models are at the other end of the spectrum. In these models, detailed mathematical descriptions of the model components and their interactions are used to reflect their actual physical and chemical behavior. Mechanism-based models enable the description and prediction of the object's behavior in great detail. But they also require very detailed and accurate information about each model component (e.g., concentrations, reaction rates etc.). And, due to the complexity of the mathematics describing the relationships between model components, mechanism-based models can only be applied for models that contain a comparatively small number of components.

The mechanistic dynamical model of metabolism that is discussed towards the end of this handout is a good representative of this class of mechanism-based model.

Constraint-based models, such as the flux-balance-analysis model discussed below, fall between these two extremes. These models do not attempt to represent the actual physical and chemical mechanisms that take place in the real-world system. Instead, they constrain the values of model parameters to stay within certain bounds established either by experimental data or by basic physical or chemical principles (e.g., conservation of mass).

The need for formal and mathematical models

For much of its history, biology has been a data-poor science. That is to say that relative to the complexity of actual biological systems, the available data have been very limited. The total number of human genes, for example, was based on vague (and as it turned out rather inaccurate) estimates until the human genome project was completed at the beginning of the millennium. Notwithstanding some very elegant mathematics-driven work in the first half of the last century (see, for example, the works of Haldane, Fisher, and Delbrück), relatively simple verbal and semi-formal models dominated biological research until recently and have been responsible for many breakthroughs.

Why then do we use formal and mathematical models? The answer to this question is twofold.

Firstly, over the last decade, biology transformed from a data-poor into a data-rich science. New high-throughput and "omics" technologies are now generating enormous amounts of information on genetic variation, gene expression as well as the abundance and interactions of proteins, nucleic acids, and metabolites. The ever-increasing size of these data sets surpasses the human ability to track or even visualize the information they contain. If you are not convinced, you can download any genome-wide data set and try to just visualize the raw data - let alone make sense of it. Only by framing these large data sets in formal or mathematical models that can be analyzed by computers can they be made tractable.

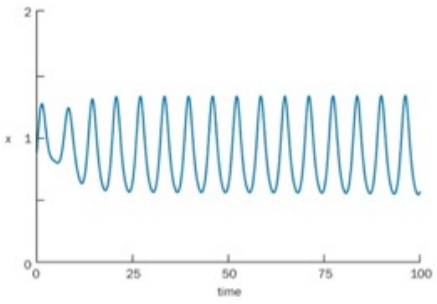
Secondly, common sense and intuition are often poor guides even for the analysis of relatively simple systems. This is particularly true for dynamic systems involving feedback loops or non-linear relationships (Figure 4).

Constraint-based models of metabolism

In the first part of the course, cellular metabolism will serve us as an example to illustrate some of the principles and methods of systems biology. Please note that these same methods and principles can also be applied to a broad range of other systems-biology questions.

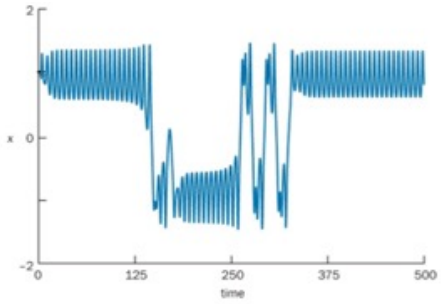
You will learn about two types of models: constraint-based (in this section) and mechanism-based models (further below).

a)

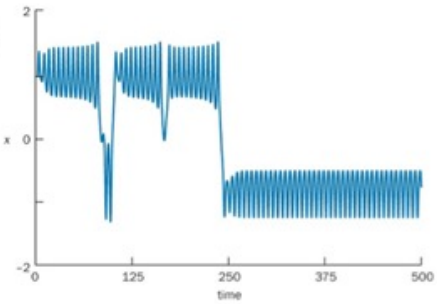


$X_0 = 0.9, \quad Y_0 = 0.4, \quad a = 0.2645$

b)

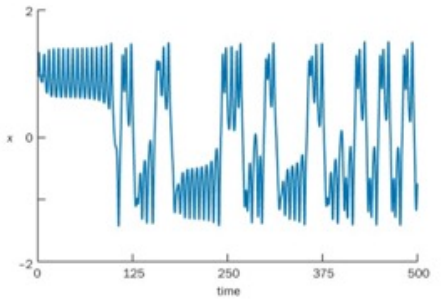


c)



$X_0 = 0.9, \quad Y_0 = 0.4, \quad a = 0.265$

d)



$X_0 = 0.91, \quad Y_0 = 0.4, \quad a = 0.2645$

e) $\frac{dX}{dt} = Y$

$\frac{dY}{dt} = X - X^3 - 0.25Y + a \sin t$

Figure 4 *The four panels of this picture show plots generated by the set of differential equations shown in panel **e).** The parameters of the equation are shown below the panels. In panel **a)**, you can see that the function initially behaves very regularly and shows the oscillation expected due to the sine term. Panel **b)** shows the same equation with the same parameters as in panel **a)** but plotted over a longer time-period. Note how at t≈125, a new and unexpected behavior of the function emerges. Panels **c)** and **d)** further show how very subtle changes in the chosen parameters can change the overall appearance of the function's time course very dramatically. Without mathematical or computational analysis, the behavior of these two equations would be very difficult to predict. (adapted from Garland Science "A First Course in Systems Biology (2013))*

Steady-state metabolism

Under conditions of a continuous nutrient supply and no overcrowding (e.g., the conditions found in a bioreactor), the metabolism of a population of microbial cells can be thought of as a steady flow of metabolites through a system of biochemical reactions. This metabolic system takes up nutrients, discards waste, and generates new biomass (e.g., in the form of new cells), but these in- and outflows are perfectly balanced such that the inside of the cell is in a steady state (Figure 5). The concentration of the metabolites inside the cells does not change over time.

This is an example of homeostasis, i.e., the tendency of living organism to keep the conditions inside their interior relatively constant.

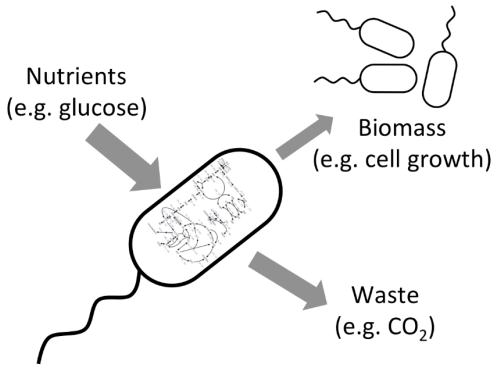


Figure 5 *Under certain growth conditions (e.g., steady supply of nutrients, no crowding), the metabolic system inside a cell exists in a steady state: The inflow of nutrients into the metabolic system is perfectly balanced by the outflow of waste and newly generated biomass (e.g., in the form of additional cells), waste, and newly generated biomass (e.g., in the form of additional cells).*

Metabolic flux vs. metabolite concentration

At first glance, it may seem that the most interesting fact to be known about a metabolite would be its concentration inside a cell. The flux through a metabolite however, is just as interesting. The word flux designates the turnover of a metabolite, which is the amount of the metabolite that is being produced and consumed over a given time-interval. So flux is the amount of biomass "flowing" through a metabolite during a certain period of time.

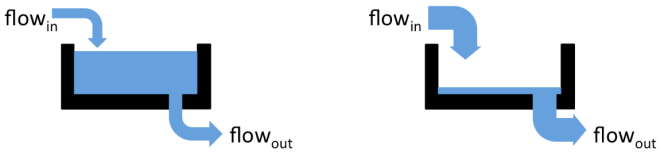


Figure 6 *Analogy between the flow of water through a pool and metabolic flux. The concentration of the metabolite corresponds to the level of water in the pool and the metabolic flux corresponds to the amount of water flowing through the pool in a given time-period. This analogy makes it clear that flux and concentration are two independent variables. The flux of a metabolite, for example, can be high, even when its concentration is low (right panel).*

In order to understand the relationship between metabolite concentration and metabolite flux, another analogy of water flowing through a pool might be helpful (Figure 6): While the concentration of the metabolite corresponds to the water level in the pool, the flux of the metabolite is the rate of the flow through the pool.

From this example, we can see that measuring metabolite concentrations does not give us information about metabolic fluxes. A particular metabolite, for instance, may be present in the cell at a high concentration even though its flux is low. This would be exactly the situation expected for a compound that the cell uses for energy storage. Conversely, we would expect metabolites in the central metabolic pathways to display a very strong flux while only being present at moderate concentrations.

Flux balance analysis

Flux balance analysis assumes a metabolic system (e.g., a cell) that is in steady state (Figure 7) and aims at understanding its inner workings in terms of the reaction fluxes.

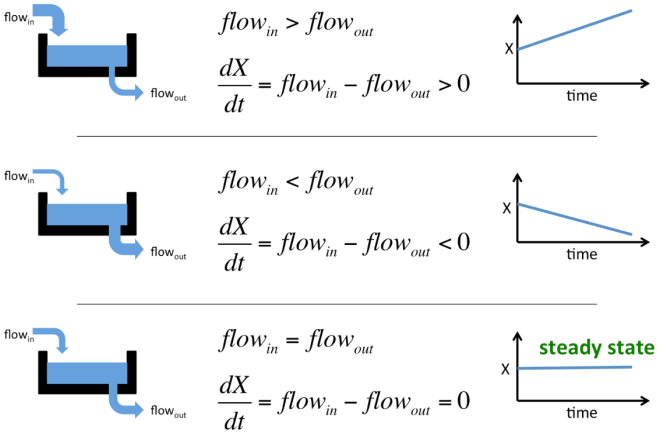


Figure 7 *Analogy between flow of a liquid and metabolic flux. The amount of a particular metabolite (X) in a cell can be thought of as a pool containing liquid. The reactions leading to the production of this metabolite can be thought of as an in-flow to the pool, while the reactions leading to a depletion of the metabolite in question can be thought of as an out-flow. If the inflow is larger than the outflow (top panel), X will accumulate over time. dX/dt represents the rate with which X changes over time (i.e., the slope of the graph on the right). Conversely, if the out-flow is larger than the inflow (middle panel), the amount of X will decrease over time. Only when the in- and out-flow are perfectly balanced will X remain the same. This situation is referred to as a steady state.*

Imagine knowing the time a system consumes by taking up and using nutrients and also knowing the time necessary for producing new biomass. You also know the stoichiometry of every enzymatic reaction inside the cell. However, neither do you know the rates of any of these enzymatic reactions nor do you know the concentration of any of the metabolites within the cell. Your goal is, nevertheless, to understand the way in which the metabolic flux is directed through your cell's metabolic network (Figure 8). Flux balance analysis is designed to solve exactly this type of problem by means of constraint-based modeling.

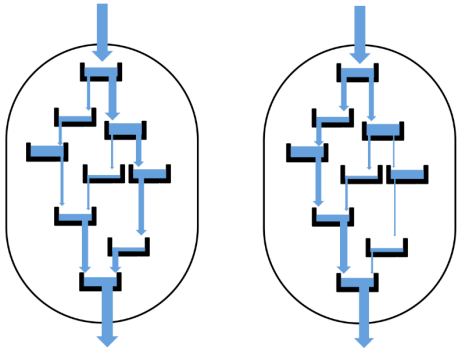


Figure 8 *Fluid analogy of the type of question addressed by flux balance analysis. Two flow systems are shown that possess the same in-flow and the same out-flow. Note that both systems are in a steady state (in flow = out flow). Also note that the water level in each pool is the same in both systems. However, the internal routing of the flow in the two networks is very different. Flux balance analysis is designed to determine the flow pattern inside steady state systems like these.*

Constraint-based modeling

"Once you eliminate the impossible, whatever remains, must be the truth." - Sherlock Holmes

Flux balance analysis is a typical example of constraint-based modeling and the strategy applied bears great resemblance to the famous quote of Sherlock Holmes shown above. One starts with a set of all possible states that a model could possibly adopt. Then constraints are applied in order to eliminate those states of the model that are inconsistent with these constraints. This typically reduces the set of possible states to a much smaller subset.

In a second step, these remaining states are evaluated according to how well they achieve a particular objective. The state of the model which best meets this demand is judged to be the best representation of the real system.

Constraints used in flux balance analysis

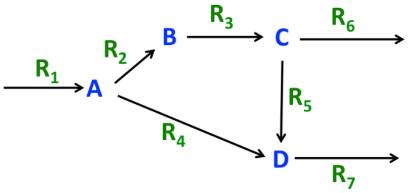
The description of the constraint-based modeling approach given in the previous paragraph is, admittedly, rather abstract. In order to get a better idea of how constraints work we will have a look at the concrete example of a simple metabolic network (Figure 9). It can be represented either as a list of reactions, as a directed graph, or as a stoichiometric matrix. In principle, the metabolic flux through these reactions could take on any value. Some reactions could be very fast (e.g., because there are many copies of the enzyme that are catalyzing this reaction) while others may be slow. *A priori,* we do not know.

But if the metabolic system is in a steady state, we know that the concentration of each of the system's metabolites has to remain constant. This can only be the case when all the fluxes in our metabolic network are balanced.

a)

Inflows:	A	Reactions:	A -> B
			A -> D
Outflows:	C		B -> C
	D		C -> D

b)



c)

$$S = \begin{array}{c} \begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & R_7 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \end{matrix}$$

Figure 9 *Three different ways of representing a metabolic flux network of four metabolites A-D (blue) and seven reactions (R_1-R_7). Panel (**a**) lists the metabolite flows into and out of the system as well as the reactions taking place inside the system. Panel (**b**) shows a graphical view of this network. Panel (**c**) shows how a metabolic flux network can be represented in a stoichiometric matrix S. Each row corresponds to a metabolite and each column to a reaction. The values in the matrix indicate the stoichiometry of the reaction. For example, column R_4 describes a reaction in which one molecule of A is converted into one molecule of D. The values in the stoichiometry matrix are always integers.*

In the case of the metabolic network in Figure 9, this means, for example, that the flux along reaction R_2 has to be exactly the same as the flux along reaction R_3. (If the flux along R_2 were greater than the flux along R_3, metabolite B would build up over time.) However, the flux along R_3 also has to be balanced with fluxes R_5 and R_6 such that the sum of the fluxes along R_5 and R_6 corresponds exactly to the flux along R_3. This implies that the flux along R_2 is also linked to the flux along R_5 and R_6, though indirectly. All the other fluxes in the network are tied together in a similar fashion.

It becomes obvious, that given the architecture of the metabolic network, only certain combinations of fluxes will result in a situation where all fluxes are balanced in a way so that the system remains in a steady state.

So the condition of a steady state introduces a large number of constraints among the different fluxes.

Mathematical formulation of a flux balance model

How can we describe these constraints in mathematical form? We shall consider the case of the simplest possible biochemical network in which metabolite X is generated by one reaction and consumed by another (Figure 10). In this case, we can describe the steady-state condition as that the inflow and the outflow are balanced, with the change of metabolite X over time being zero. Here, we can quantify the flow of the metabolite by a stoichiometric coefficient and a flux variable. We can think of the stoichiometric coefficient as the number of molecules of X that are produced (or consumed) when the reaction in question takes place once. And the flux variable v indicates how often that reaction takes place.

More generally, we can describe this particular steady-state condition of a metabolite X as

$$\frac{dX}{dt} = s_1 \cdot v_1 + s_2 \cdot v_2 + s_3 \cdot v_3 \ldots = 0,$$

where each term $s_i \cdot v_i$ corresponds to the amount of metabolite X contributed by a reaction R_i. Again, the stoichiometry coefficient s_i indicates how many molecules of X are generated (or consumed, if s_i is negative) each time reaction R_i takes place. And the flux variable v_i indicates how often this reaction takes place.

Note that with the help of a stoichiometry matrix we can set up these types of equations very rapidly for all the metabolites in a metabolic network. In the case of the reaction network shown in Figure 9, we can write down the steady-state condition for metabolite A as:

$$\frac{dA}{dt} = 1 \cdot v_1 + (-1) \cdot v_2 + 0 \cdot v_3 + (-1) \cdot v_4 + 0 \cdot v_5 + 0 \cdot v_6 + 0 \cdot v_7 = 0$$

the stoichiometry coefficients in this equation being the values from row A of the stoichiometry matrix (Figure 9c).

We can proceed in a similar manner for all other metabolites in our network, obtaining the following system of equations:

$$\frac{dA}{dt} = 1 \cdot v_1 + (-1) \cdot v_2 + 0 \cdot v_3 + (-1) \cdot v_4 + 0 \cdot v_5 + 0 \cdot v_6 + 0 \cdot v_7 = 0$$

$$\frac{dB}{dt} = 0 \cdot v_1 + 1 \cdot v_2 + (-1) \cdot v_3 + 0 \cdot v_4 + 0 \cdot v_5 + 0 \cdot v_6 + 0 \cdot v_7 = 0$$

$$\frac{dC}{dt} = 0 \cdot v_1 + 0 \cdot v_2 + 1 \cdot v_3 + 0 \cdot v_4 + (-1) \cdot v_5 + (-1) \cdot v_6 + 0 \cdot v_7 = 0$$

$$\frac{dD}{dt} = 0 \cdot v_1 + 0 \cdot v_2 + 0 \cdot v_3 + 1 \cdot v_4 + 1 \cdot v_5 + 0 \cdot v_6 + (-1) \cdot v_7 = 0$$

It is important to note that these equations are not independent of one another, but form a system. This means that the value for a particular flux variable v_i has to be the same in all four equations. A set of v values (v_1, v_2, v_3, v_4, v_5, v_6, v_7) for which these four equations are fulfilled represents a combination of fluxes that will maintain our metabolic network in a steady state.

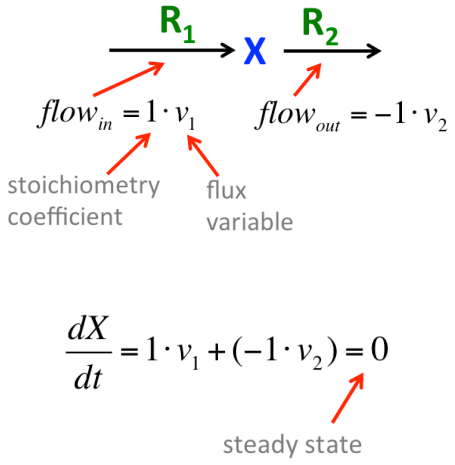


Figure 10 *Mathematical formulation of a steady-state condition explained with the help of the simplest possible metabolic network.*

You may have noticed, that this system of equations consists of four equations, even though there are seven variables (v_1-v_7). This means that our system is not fully determined but only constrained. There will not be one unique set of *v* values that fulfills these equations. Instead, there will be many valid sets of *v*-values and each one of these sets represents a possible flux pattern that will keep our metabolic network in a steady state.

Matrix representation of a flux balance model

For real metabolic networks, the systems of equations are, of course, far more complex than our example above. By using matrix notation, we can represent them in a much neater way (Figure 11). The system of equations we depicted above can easily be formulated as a matrix-times-vector multiplication. The flux vectors ***v*** that fulfill this equation constitute the flux patterns that will keep the metabolic network described by matrix ***S*** in a steady state.

If the equivalence of these two notations is not immediately clear, the following link to a short video on multiplying matrices with column vectors might be helpful.

https://www.khanacademy.org/math/precalculus/precalc-matrices/matrix_multiplication/v/multiplying-a-matrix-by-a-column-vector

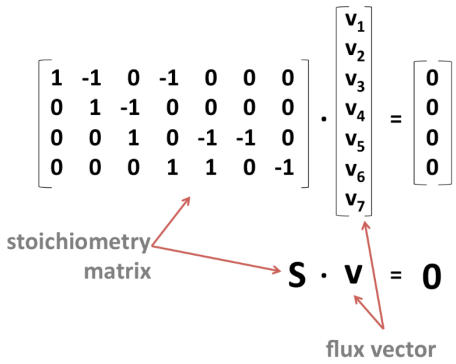


Figure 11 *Matrix representation of the flux balance model for the metabolic network shown in Figure 9. Given the stoichiometry matrix S for a metabolic network we are looking for a set of fluxes (i.e., a flux vector v), for which the product of the stoichiometry matrix and the flux vector is a null vector. For this set of fluxes, the entire metabolic network will be in a steady state.*

Additional constraints can be incorporated in the matrix model

In addition to the constraints introduced by the stoichiometry of the metabolic network metabolic fluxes are also constrained by other factors, such as the maximal rate of the take-up or release of a metabolite or an upper or lower limit on a specific metabolic flux.

The objective function: Biological systems have evolved to optimize specific objectives

We now have at our disposal a very compact way of formulating the constraints of our flux balance model as a matrix-times-vector equation. We have also seen that this equation does not necessarily give us one single flux vector as a solution, but rather gives us a large set of possible flux vectors instead. Each of these solutions will keep our metabolic system in a steady state, but the actual flux patterns represented by them may be very different physiologically. How do we determine which of these flux patterns is the one that best reflects the situation in the actual metabolic system that we are modeling?

This is where the objective function comes in. Biological systems distinguish themselves from other systems by the fact that they have been shaped by evolution. They have been optimized to achieve particular objectives that help an organism to survive and compete with other organisms. Examples could be a maximal growth rate, maximal efficiency in the use of available resources, or an optimal resistance to drought.

We can conclude, that an organism which can achieve a steady-state metabolism with many different flux patterns (i.e., flux vectors) will choose the pattern which best meets its objectives.

The objective function is a function $f(v_i, v_{2,}\dots)$ that calculates a numerical value representing how well a particular flux vector will achieve the organism's objective(s). For example, if the survival of the hypothetical organism, whose metabolism is represented in Figure 9, would depend primarily on secreting metabolite D, the objective function would simply be $f = -v_7 \cdot s_{4,7}$, which is the rate at which this metabolite is released.

For many microorganisms it seems reasonable to assume that the objective is to outcompete other organisms by growing as fast as possible. This is particularly true considering the growth conditions (plentiful and steady nutrient supply, no crowding) under which flux balance analysis is often done. In terms of metabolism, maximal growth corresponds to the fastest possible generation of the amino-acids, lipids, nucleic acids etc. needed to generate biomass (i.e., new cells). The objective function for maximal growth would then be the sum of the metabolic fluxes that generate these biomass precursor molecules.

While this example of an objective function seems to be reasonable, defining sensible objective functions that reflect the actual biological challenges faced by organisms still remains an open research question. You

will learn more about different objective functions later on in the course.

Linear programming: Finding the flux vector that meets all constraints and optimizes the objective function

Our problem now is to find a mathematical method that finds all flux vectors that meet the constraints set up by our stoichiometry matrix and to select from these vectors one that maximizes the objective function.

This may seem like a difficult task, but fortunately, this type of problem normally (see below) corresponds to a well-known class of general mathematical problems known as linear programming (or linear optimization) problems.

Linear-programming problems occur in many fields (finance, traffic planning, logistics etc.) and have been studied extensively for more than half a century. Some very efficient algorithms are available for solving them, e.g., the simplex algorithm. (https://en.wikipedia.org/wiki/Simplex_algorithm)

The term "linear" in linear programming refers to the fact that both the equations setting up the constraints as well as the equations for calculating the objective function must be linear equations. If any of these equations are non-linear, the linear-programming algorithms cannot be used. Other optimization methods can be used in such cases, but they tend to be computationally much less efficient. Luckily, the constraints represented in our stoichiometry matrix are, by definition, linear. The objective function can usually also be expressed in linear terms without too much difficulty.

If you want to understand how linear programming works in detail, please have a look at the following video, which you can also find in this week's section on Moodle. In the video, a simple linear programming problem is solved step by step. https://youtu.be/HE9_6qiHjhg

Dynamical models of metabolism

"The only thing constant in life is change." -

François de la Rochefoucauld

Motivation for using dynamical models

Steady-state models, such as the flux balance analysis models we discussed above, are very efficient in describing biological systems that are in a steady state.

However, biological systems and their environments are rarely in a steady state. Instead, they tend to be in a state of permanent change, due, for instance, to them having to respond to environmental changes.

Dynamical models have been designed to formally describe changes of a non-steady-state system over time.

To better understand the difference between models assuming steady-state (such as flux balance analysis) and dynamical models, consider the example of a microbe that can grow on two different carbon sources (e.g., fructose and glucose). Using flux balance analysis we could model the metabolism of the microbe on either of these carbon sources and compare how the metabolism differs under these two conditions. But, neither will steady-state models ever tell us how long it takes for the microbe to adapt its metabolism to the new carbon source, nor will they let us see, if, for instance, an abrupt switch in the carbon source leads to the production of a toxic intermediate that could kill the organism. To answer this type of questions we need dynamical models.

Other examples of questions that are typically addressed by dynamical models are: How long does it take a biochemical signal to travel through a signal transduction cascade? Or, what time course for delivering chemotherapy medications is ideal if the maximal therapeutic benefit is to be achieved with minimal side effects? Is a single large dose of the drug or a continuous infusion the better choice?

Ordinary differential equations (ODEs)

While you will undoubtedly have encountered ordinary differential equations before, this type of equation is so central to dynamical models that a short review is presented here.

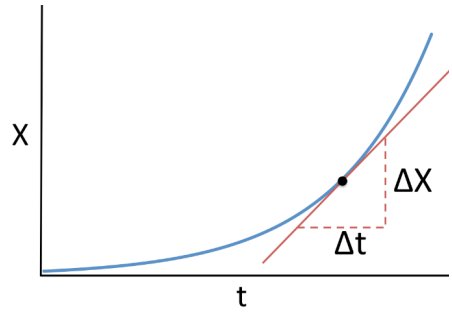


Figure 12 *Depicted in blue is a curve showing how the variable X (e.g., the concentration of a metabolite) increases exponentially as a function of time:* $\boldsymbol{X_t = e^{kt}}$*. In dynamical modeling, we typically do not start with this function but with information about the rate with which X changes along this curve. This rate corresponds to the slope and is shown in red for one point of the curve. In this case, the formula for the rate is* $\boldsymbol{dx/dt = k \cdot X}$*. From this, we can reconstruct the underlying function (blue). In the case of exponential growth as shown here, the slope of the blue graph depends on the value of X itself. Note how the slope of the blue line increases with increasing values of X.*

ODEs are equations that describe how quickly one variable changes as a function of another variable. In the case of dynamical systems, we are specifically dealing with ODEs that describe the rate with which a variable changes over time. Thus, the ODEs describing metabolic systems are sometimes called rate equations.

One of the central tasks of dynamical modeling is to "convert" information about the rate with which X changes into a function that describes how X changes as a function of time. This point is illustrated in Figure 12.

In the case of simple ODE's, an explicit analytical solution can frequently be found "by hand". For example, for the ODE

$$\frac{dX}{dt} = k \cdot X$$

the general solution has the form

$$X_t = A \cdot e^{kt}$$

where A is the value of X at *t*=0.

The analytical solution of an ODE allows us to calculate the value of X at any desired time, which puts us in a position to plot the time-course of X for any given starting value of X.

Numerical solutions to ODEs

As you will see later on, the ODE models describing metabolism will typically involve ODEs that are vastly more complicated than the example given above. Also, we will usually be dealing not with single ODEs, but with systems of multiple ODEs, where the solutions sought after have to fulfill all the ODEs in the system. They can rarely be found analytically.

Instead, complicated systems of ODEs are typically solved using numerical analysis techniques (Figure 13). The software performing the numerical analysis starts with a set of metabolite concentrations X_{t_0} that are present at time t_0. The rate equations are then used to extrapolate the concentration of X after a small time step Δt. At the techniques' most simple, the algorithm uses the slope at point X and extrapolates the value of $X_{t_0+\Delta t}$ linearly.

The resulting metabolite concentrations $X_{t_0+\Delta t}$ form the basis for the next step. Following this approach, we obtain a time series with the values for X that, when done properly, recapitulates the true time course of X. Of course, strictly speaking, this stepwise procedure is only correct if the time steps are infinitely small (i.e., $\Delta t = dt$). The smaller the time steps (Δt), however, the more steps and the more computing time are required for the analysis. The art of numerical analysis therefore lies partly in selecting step sizes that are large enough to allow the analysis to proceed at a reasonable speed, but small enough to yield accurate results.

Fortunately, numerical analysis of ODEs is a problem that occurs in many fields including engineering, physics, and economics. As a result, many very effective algorithms have been developed in order to solve it. These algorithms not only adapt step sizes automatically but also use methods to find the best linear approximations for a given section of the curve.

Mechanistic dynamical models of metabolism

Mechanistic models of metabolism are conceptually very simple. For each metabolite, we consider all the reactions that generate this metabolite and all the reactions that consume it. If we know the rates of these reactions, we can setup a set of ODEs, solve them, and thus obtain continuous time traces for each metabolite as well as the flux of biomass through each metabolite.

It is important to realize that the rate of a reaction is typically not constant but rather depends on a large number of parameters, such as the concentration of the enzyme that catalyzes the reaction, the concentration and spatial distribution of the substrate and product, the concentration of cofactors or allosteric regulators, the temperature, pH, and so on.

If we were to take all of these factors into account for every reaction, the ODEs would become extremely complicated. We would also have to obtain access to lots of information on the initial conditions inside the cell.

So the challenge in setting up a good mechanistic model of metabolism lies in finding a description that captures

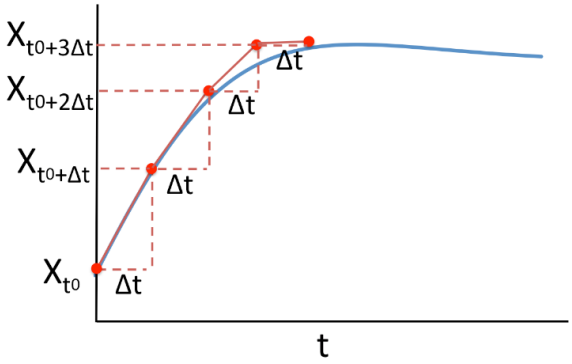


Figure 13 *Numerical methods for solving ODEs approximate the true function of X (blue) by short linear segments (red). In its simplest form, a numerical analysis calculates these segments by using the rate equations to estimate the slope at a given value of X and linearly extrapolates that slope over a time of Δt. For functions that vary smoothly and steadily over time, this approximation can be very accurate, even for large Δt. But when the curvature is more pronounced, the approximation quickly becomes inaccurate.*

the essence of the biochemical mechanism in a mathematically compact form.

The Michaelis-Menten equation as a model for enzyme-mediated reactions

The Michaelis-Menten equation is widely used to model enzymatic reactions. You will certainly have encountered it during the course of your studies. The description here is therefore merely intended as a refresher and is kept quite concise.

For the simple enzymatic reaction shown in Figure 14, we can easily formulate ODEs (see below) to describe the rates with which the concentration of each molecular species in this reaction changes over time.

Formulating these equations is straightforward (as an exercise, you may set up these equations yourself), but

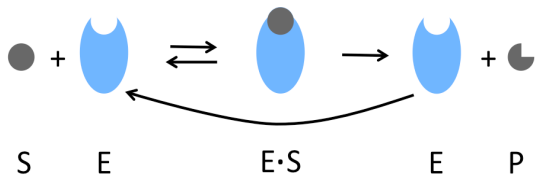


Figure 14 *Schematic representation of a reaction in which an enzyme (E) catalyzes the conversion of a substrate (S) into a product (P) via an intermediate enzyme-substrate complex (E•S).*

you will notice that all four equations including those for the consumption of the substrate and the production of the product contain a term involving the concentration of the enzyme-substrate complex. Measuring this concentration experimentally or calculating it from experimentally measurable parameters is exceedingly difficult. The ODEs, though mathematically correct, turn out to be of limited practical value for computing the rate of substrate-to-product conversion in this reaction.

$$\frac{d[S]}{dt} = -k_1[E][S] + k_{-1}[E \cdot S]$$

$$\frac{d[E]}{dt} = -k_1[E][S] + k_{-1}[E \cdot S] + k_2[E \cdot S]$$

$$\frac{d[E \cdot S]}{dt} = k_1[E][S] - k_{-1}[E \cdot S] - k_2[E \cdot S]$$

$$\frac{d[P]}{dt} = k_2[E \cdot S]$$

However, by assuming that the concentration of the enzyme substrate complex stays the same throughout the course of the reaction (this is called the quasi-steady state assumption), the equation for the production of the product can be transformed easily into the famous Michaelis-Menten equation

$$\frac{d[P]}{dt} = v = \frac{V_{\max}[S]}{k_M + [S]}$$

where $k_M = \frac{k_{-1}+k_2}{k_1}$

and $v_{\max} = k_2([E] + [ES])$.

The major advantage is that its parameters can be measured in a relatively straightforward fashion. [E]+[ES] is the total concentration of the enzyme, which can be measured relatively easily (e.g., via mass spectroscopy). k_2 and k_M are properties of the enzyme that can be measured experimentally *in vitro* and that have been cataloged for a very large number of enzymes and are available in online data bases.