

Introduction to Modelling in Systems Biology

Systems Biology Models (8BM050)

M. de Rooij

version: 08-2024



Contents

1 Introduction to Systems Biology	3
1.1 Studying Biological Processes as Interconnected Systems	3
1.2 Models and Modelling	4
1.3 Outline of the Course	5
1.4 Mathematics and Programming in Modelling	5
1.5 Digital Twins	6
2 Network Models	7
2.1 Introduction to Graph Theory	7
2.2 Metabolic Networks	14
2.3 Exercises	16
3 Dynamic Models	17
3.1 Biochemical Systems	17
3.2 Signaling Systems	24
3.3 Steady-States, Setpoints, and Perturbations	33
3.4 Measurements	35
3.5 Pharmacokinetic Modeling	36
3.6 Exercises	41
Bibliography	42

Introduction to Systems Biology

1

"The problem is not the problem. The problem is your attitude about the problem."

— Jack Sparrow (Pirates of the Caribbean)

1.1 | Studying Biological Processes as Interconnected Systems

For the greater part of the 20th century, biology was dominated by the so-called *reductionist* view. The central idea was that by detailed examination of each component in the system, we would gain an understanding of the system as a whole. While the obtained knowledge from reductionist research is useful in studying biological systems, reducing these to the sum of their parts has disadvantages that relate to some key properties in complex regulatory systems.

The first of these is *emergence*. Complex systems regularly display properties related to the combined interaction of the components. These properties cannot be directly related to the components alone. This is often observed in chemistry, as knowledge about the interaction between atoms is not enough to understand the interaction between molecules, as they display emergent properties in specific configurations, which we need to understand as well.

A second property is *redundancy*. The inherent robustness of biological systems to perturbations is what keeps us alive. Biological systems are robust because of a large redundancy in their components. This means that different components can compensate together in situations when a component is lacking.

The third and last key property is *modularity*. As also seen in various other courses on biology, our body can be studied on multiple levels, as it is composed of various organ systems, that are each made up of organs. Each organ in our body is built from forms of tissue, which is made up of cells. This strongly hierarchical property requires us to study processes on multiple levels. [1]

A common way to gain understanding about these systems as a whole is through combination of experimental 'wet lab' research (*in vitro* and *in vivo* experiments), and computational (*in silico*) modelling. The latter of which is an important component in systems biology. This field is largely concerned with the study of interacting biological components, not by merely investigating the components individually, but by examining the living system as a whole. A common example to illustrate the utility of the approach taken in systems biology can be seen in Figure 1. In this figure, six men are each investigating a part of the whole system (elephant), while only by studying the system as a whole, the correct solution can be found.

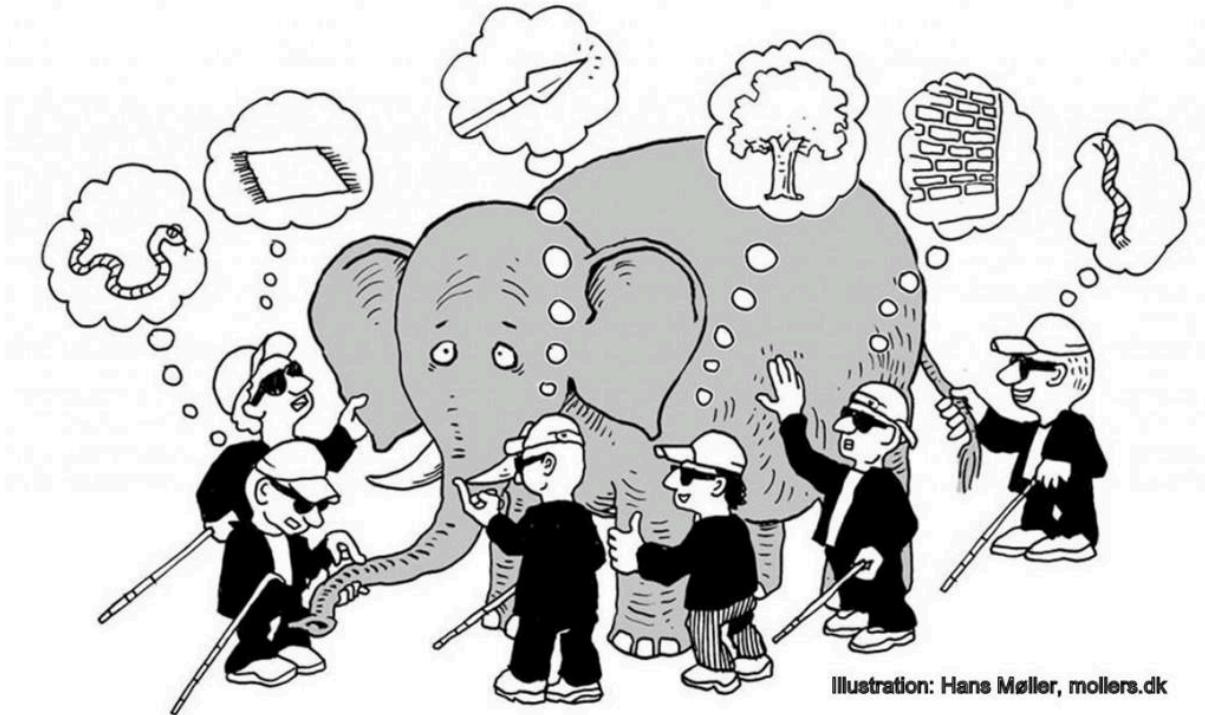


Illustration: Hans Møller, mollers.dk

Figure 1 : Six blind men investigating parts of the elephant. Each of the men has a different hypothesis of what the elephant could be, based on them feeling a specific part of the elephant. However, only by investigating all the specific parts, they can conclude that there is an elephant standing before them.

This practice can be applied to study mechanisms that keep us alive, such as homeostatic systems, or cellular communication networks, and to identify causes of disease and the effect of possible treatments. [2] In these lecture notes, we aim to provide an introduction to integrating biological knowledge and findings from wet lab research into models, analyzing these models, and identifying basic properties of these systems that are otherwise difficult or impossible to study in practical experiments alone. In this way, we hope to create an understanding of the utility and importance of models as a tool to aid engineering research. [3]

1.2 | Models and Modelling

Before we can talk about different modelling techniques, it is important to clarify what we mean with modelling and models in systems biology. A model in science and engineering is a broad concept, but is always something that is made to reflect a part of a real system, often in a simple, more malleable way, in order to study specific components of that system. We can do this by constructing a physical system, such as a pump and a set of tubes to model circulation, through a living model, such as a mouse model, or by means of a computer model, often described in some mathematical framework. These models are all experimental setups, but we can also have conceptual models, such as a diagram of different interactions in a system, or a free body diagram in mechanics. An example of an interaction diagram can be seen in Figure 2.

In this course, we will be looking at computer models. However, we will only be discussing a very specific type of model. While systems biology is a broad field, we are not diving into

typical models that are used for machine learning, such as neural networks, or linear regression.

Instead, in this course, when we talk about a model, we will be referring to a simplified mathematical description of a biological process. The mathematical framework for each model can vary depending on the level of detail we wish to include, the amount of information we possess, and the questions we wish to answer using the model.

Explain more in depth what modelling is; also include bottom-up vs top-down

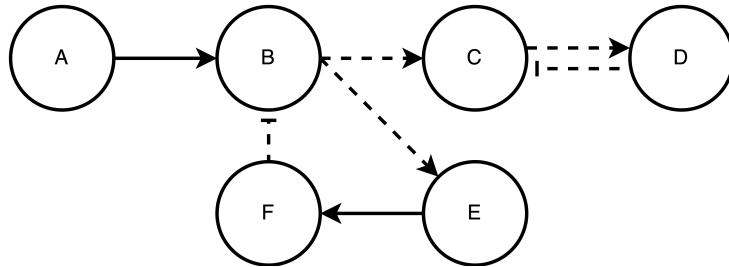


Figure 2 : Example of a conceptual model of an interaction system between various components. In this arrows present stimulatory interactions whereas bars indicate inhibitory processes. The solid lines depict the conversion of one component into another, and the dashed lines represent communicative interactions (where the messenger is not consumed). Models like these are commonly used to show the structure of various mathematical models.

1.3 | Outline of the Course

Over the course, we will be highlighting several types of mathematical models. We will gradually introduce new concepts that arise from the various mathematical frameworks that these models are embedded in. Initially, we will describe how to combine sources of biological evidence to form hypotheses about associations between components. These hypotheses can be combined to form an associative model of interactions. While these are useful in grasping the general function of the system, these models provide limited ways of analyzing the underlying system. Mathematical graph theory can aid us in obtaining static properties of systems and to identify the important components. Therefore, we will shortly discuss the fundamentals of this branch of mathematics, and how this connects to biological meaning. Beyond static graphs, we will move into the world of dynamic modelling and explore how to create simplified models that can describe behavior of a system over time. Finally, we will touch upon what happens when we scale-up these systems to incorporate much more interactions, and how to deal with difficulties that arise as a result. Besides the mathematical frameworks, we will discuss use cases of these models, and how they can aid us in forming and validating hypotheses, designing targeted experiments, and drawing comprehensive conclusions.

The goal of this course is therefore to provide its students with a strong set of tools that allow them to begin modelling biological systems on a great range of scales and in various medical contexts. But also to give them the ability to use these models as an aid in future experimental research.

1.4 | Mathematics and Programming in Modelling

Within these lecture notes, various concepts of mathematics will be treated. Some concepts, like differential equations, you will have seen in previous courses. Others, such as

graph theory, may be new. It is important to highlight that we view mathematics as a method of describing our models and subsequent analyses. While the subject may be interesting and generally important, this course will not touch upon mathematical proofs underlying the various concepts. The description of the mathematics in these lecture notes is purely from a practical point of view. Even though we will need mathematical notation, the equations and notations are made to be as simple and easy-to-read as possible. Furthermore, all equations are explained in as much detail as necessary for understanding their use.

Besides mathematics, this course will also feature some programming skills. As with mathematics, these programming languages are tools that we use to perform simulations and analyses. It is not our goal to make you write the most beautiful and efficient programs that exist. What is important is that after the course, you will be able to use appropriate software in modelling. In further parts of these lecture notes, modelling and analyses will be accompanied by examples written in Python, to illustrate how you could perform various procedures yourself.

1.5 | Digital Twins

Introduction to the concept of digital twins

Network Models

2

"If you don't know where you want to go, then it doesn't matter which path you take."

— The Cheshire Cat (Alice in Wonderland)

2.1 | Introduction to Graph Theory

In this section, we will introduce you to the fundamentals of mathematical graph theory, and apply this to analyze biological systems. Of course, we first will have to explain what a graph is. But first, one often encountered example to illustrate the significance of graph theory.

Example 1

The city of Königsberg - now called Kaliningrad - was located on the river Pregel in what was called Prussia back then. The city contained two islands in the river, and two banks on each side of the river. The regions were connected by seven bridges as shown in Figure 3. The problem in this case was to find a path through the city by crossing each bridge exactly once.

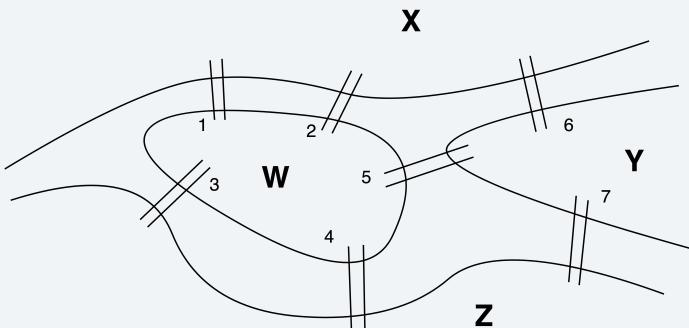


Figure 3 : The Bridges of Königsberg as a drawn map.

Famous mathematician Euler pointed out that the only important property in this case, was the sequence of bridges taken, and that the paths taken on each land mass are irrelevant. This led to Euler being able to simplify the problem into a set of points - representing each land mass - and lines - representing each bridge connecting two land masses - in one figure. The result is shown in Figure 4.

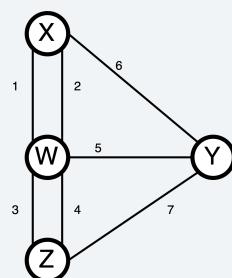


Figure 4 : The Bridges of Königsberg as a graph.

He devised that it was *impossible* to find a path that crossed every bridge, but each bridge only once, as the fundamental requirement for this to be possible, was that either zero, or two of the points had an uneven number of lines connected to them. Since there are four points in Figure 4 with an odd number of lines connected to them, such a path is not possible. Now, this argument is considered to be the first theorem of graph theory, and a path that traverses every line once, is called an Eulerian path. [4]

Slightly more formally, a graph is a set of points, called ‘vertices’ or ‘nodes’, connected by lines, called ‘edges’. We generally denote this by using the letters G for a graph, V for vertices and E for edges in between accolades. Therefore, we can note down any graph as

$$G = \{V, E\} \quad (1)$$

If our graph is small enough, we can often visualize it easily, as was the case with the example above. Another example visualization of such a small graph can be seen in Figure 5. This graph contains eight vertices, labelled using the letters A-H, and nine edges. Besides describing graphs this way, we can also use what we call an adjacency matrix. This is a square matrix of size $|V|$, which means that it has the same number of rows as it has columns, which is equal to the number of vertices in the graph. As you may have already deduced from the notation, we use bars around the letters V and E to denote the numbers of vertices and edges respectively.

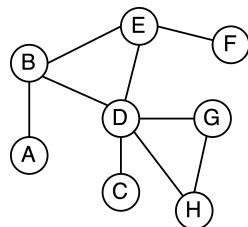


Figure 5 : An example visualization of a simple graph, with 8 vertices and 9 edges. For small graphs, we can easily make these figures, but for larger graphs, it quickly becomes difficult to prevent overlapping edges, and to distinguish connections.

In its simplest form, the adjacency matrix is both square and *symmetric*, which means that on coordinate $[i, j]$ you will find the exact same entry as on coordinate $[j, i]$. Moreover, you will find only the numbers 0 and 1 in this matrix, where a 0 in position $[i, j]$ denotes that two vertices, represented by row i and column j are *unconnected* and a 1 in this position indicates that there is a connection between these two vertices. We can form the adjacency matrix for the graph in Figure 5 as follows

$$\left(\begin{array}{ccccccccc|c} A & B & C & D & E & F & G & H & \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & | & A \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & | & B \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & | & C \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & | & D \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & | & E \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & | & F \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & | & G \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & | & H \end{array} \right) \quad (2)$$

Another thing you might notice is that the diagonal of this matrix contains only zeros. This is because this simple graph does not contain *self-loops*, which are connections of one node to itself. In case of a self-loop, the entry 2 is used in the adjacency matrix.

In a computer, a graph is usually represented by an adjacency matrix, because it is useful to store large amounts of data in these matrices. Also, when performing analyses on graphs, these matrices come in handy, because we can apply some operations from linear algebra on these graphs to quickly and efficiently get our information out. An example calculation you can do is to sum each row, or column, to obtain the amount of connections one vertex has in total. A more complicated operation is to multiply one column with the transpose of a row. This will provide you with the amount of connections two vertices have in common. These and more of these properties will be discussed in a later section in more detail.

2.1.1 | Types of Graphs

We can distinguish types of graphs, that represent connections between vertices in different ways. The simplest type is the one seen above, which is called an *undirected graph*, which means that all edges do not represent a direction in the connection between the vertices. Instead of this, we can also provide a directionality to edges, giving us a *directed graph*. In these graphs, the adjacency matrix is not symmetric anymore, because connecting a vertex *A* to vertex *B*, does not directly mean that vertex *B* is also connected to vertex *A*. Biological signals or chemical reactions often have distinct directions, which also means that we often represent those with directed graphs. Besides directionality, we can distinguish graphs based on whether they have *self-loops*, which are vertices that are connected to themselves, or whether they are *cyclic* (you can draw a path from a vertex that ends in the same vertex, without using the same edge twice), or the opposite: *acyclic*. Examples of these different types are shown in Figure 6.

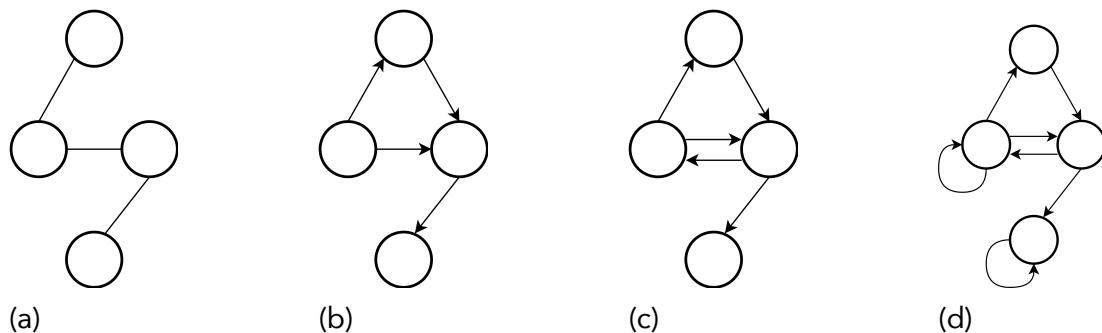


Figure 6 : Different graph types. **(a)** An undirected acyclic graph. The edges have no directionality and there are no cycles. **(b)** A directed acyclic graph. The edges have a directionality, but the graph is not cyclic. **(c)** A directed cyclic graph, where we have directionality and cycles. **(d)** A directed cyclic graph with self-loops.

It may be difficult to make this information more concrete right now, so let's link these graph types to biological systems so we can understand their different uses. Because many systems in biology are linked to directions, the undirected graph may be difficult to put to use directly.

An example where we can use undirected graphs is when analyzing gene expression data. For example, we may want to analyze whether two genes are co-expressed, meaning that mRNA is often produced from two genes at the same time of measurement [5]. We can construct such a graph by creating a vertex for each gene, and connecting two vertices if

the expression of these genes is *significantly* correlated (testing if it is true that when gene a has a high expression value, gene b also is likely to have a high expression value). In that way, if we know what the function of gene a is, we can hypothesize about the possibly yet unknown function of gene b .

For a directed graph, applications within biology are possibly easier to think of. Of course, we can create a directed graph of chemical reactions, or pathways, such as glycolysis, or the MAP kinase pathway. Using these graphs, we can easily reason about the downstream effect of small upstream changes. Self-loops are also often observed in biological systems, such as where a piece of DNA encodes for a protein that blocks itself from being transcribed, creating a feedback loop.

2.1.2 | Graphs in Biology

Besides the coupling of names, letters or numbers to vertices, we can also add other properties of our system to vertices. We could, for example, combine enzymes and reactants and products together in one directed graph, where we use the directionality to distinguish between the product and the reactant in a catalyzed reaction. In this graph, we will have an additional vertex property; whether it corresponds to a (consumed or produced) molecule, or an enzyme. An example is shown in Figure 7, showing the conversion of glucose to glucose-6-phosphate, by a hexokinase enzyme. Using this graph, we can identify whether a type of cell missing a specific enzyme would still be able to produce certain molecules. Other properties we could add to the enzyme nodes are the Michaelis-Menten constants (K_M) for these specific enzymes.

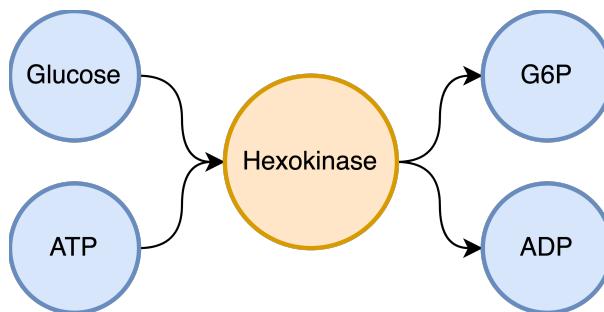


Figure 7 : Example of a directed graph with two types of vertices using the conversion of glucose to glucose-6-phosphate (G6P) upon entry into the cytoplasm. We have the enzyme vertex(orange), and the four molecule vertices (blue).

When talking about chemical reactions, we can also think about properties we could add to edges. In a simple chemical reaction graph, we can add the rate at which the reaction happens (k) as a property to the edge. Therefore, besides the structural component of graphs, we can also add data to each graph element. Because biological systems tend to be incredibly complicated, graphs are a powerful tool to analyze these systems in a comprehensive way. In the next section, we will discuss some graph properties that can aid us in analyzing a biological system.

2.1.3 | Properties of Graphs

A graph is not just created to make visualizations of connectivity within a system. After constructing a graph, we often also want to obtain properties of this graph, and reason about what these properties mean for our system in real life. We can distinguish properties

at the graph level, and at the vertex and edge level. We will first discuss some basic properties of graphs and their elements.

Degree

The degree of a vertex in a graph is the number of edges that connects it to the rest of the graph. In directed graphs, we can distinguish an in-degree and an out-degree for incoming and outgoing edges, respectively. The degree of a vertex is difficult to interpret directly on its own, as its meaning strongly depends on the degree of all vertices in a graph. An important measure of a graph, however, is the *degree distribution*. It is given by the distribution $P(k)$ for degree k as

$$P(k) = \frac{n_k}{n} \quad (3)$$

Where n_k is the number of vertices having degree k and n the total number of vertices. The sum of all degrees in a graph is directly related to the number of edges by

$$\sum_n k_n = 2|E| \quad (4)$$

In words, this says that the sum of all degrees in a graph equals twice the amount of edges. Therefore, the sum of all degrees in a graph must always be even.

Scale-Free Graphs

A popular property related to degree distributions and often attributed to metabolic networks is scale-freeness. [6] The main observation in a scale-free network is that the degree-distribution follows a power-law, which is given by

$$P_{\text{sf}}(k) = k^{-\gamma} \quad (5)$$

Where the parameter γ typically lies between a value of 2 and 3.

This property relates to the idea that in a large network, there are a few vertices that are connected to many others, while there are many vertices that are only connected to one or a few others. Vertices that have many connections in such a network are called *hubs*. A simple analogy for understanding the property is by imagining a graph of all people in the world, where two people are connected if they've shaken hands. You can imagine that there are a few people, such as the president of the United States, that have shaken hands with many people, while many people have only shaken hands with a couple hundred people from their neighborhood. Similarly, on a social media platform like Instagram, there are many people where the sum of followers and following is lower than 1000, while only a few people have a sum of followers and following of more than a million.

While it was initially thought that all of these networks were scale-free, studies have shown that while they sometimes approach scale-freeness, very few of them actually strongly adhere to the power law property.

Connectedness

A basic property that most graphs you'll encounter have, is connectedness. An undirected graph is connected if we can reach every other vertex in the graph, starting from one random vertex. Considering directed graphs, we can distinguish various levels of connectedness. We call a directed graph *weakly connected*, if it becomes a connected

graph when we do not consider the directionality of the edges. In the case where there is either a path from vertex a to vertex b , or the other way around, for every vertex in the graph, we can call this graph semi-connected. When for every vertex a and b in the graph, we have both a path from a to b , as well as a path from b to a , we call this graph strongly connected.

Walks, Trails and Paths

When investigating graphs, we may also be interested in the ways that we can reach one vertex from the other. In order to properly define properties that relate to this concept, we identify three ways in which we can represent this. (1) A walk is a sequence of vertices, where each next vertex is connected by the previous. (2) A trail is a walk, but with the extra requirement that each edge is distinct. (3) A path is a trail (so, all edges are distinct) where all vertices are distinct. Additionally, a cycle is a path that has the same starting and ending vertex.

A quantity that is often of importance in a graph is the length of the shortest path from vertex A to vertex B. A famous algorithm for computing this is Dijkstra's algorithm. The details of this algorithm are beyond the scope of this course, but its implementation exists in the networkx Python package, where it can be used to find shortest paths and their lengths in graphs. An example of the shortest paths in a graph can be seen in Figure 8b.

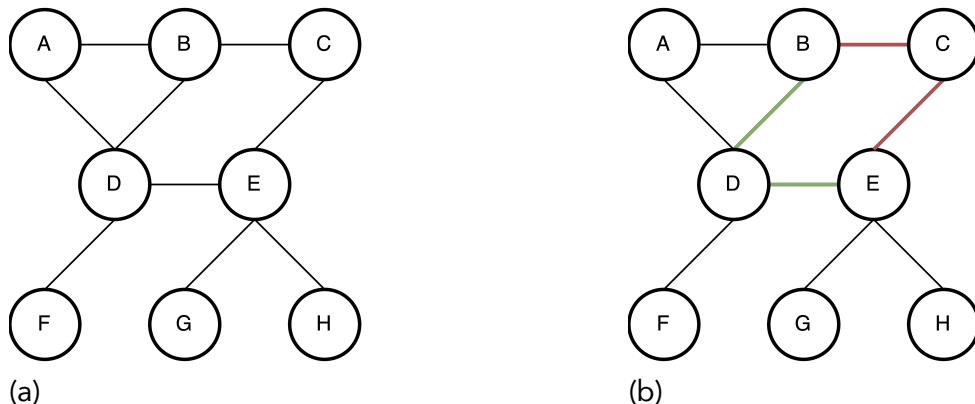


Figure 8 : (a) Example graph. Examples of walks include [B-A-B-D], [B-D-B-C], [E-C-B-D-E-H], [G-E-F]. The last two are also both walks, because we can use distinct edges. Only the last one is a path, as it contains no distinct edges or vertices. (b) Shortest paths from B to E marked in green and red. Both have length 2.

Many other properties relate in some way to paths or distances. For a vertex A, we can compute its eccentricity as the maximum distance of the shortest paths from A to all other vertices in the graph. The radius of a graph is then related to this, as this is defined by the minimum eccentricity of all vertices in the graph, and the diameter of a graph is the maximum eccentricity of the vertices in a graph. Furthermore, if a vertex has an eccentricity which is equal to the radius of the graph, this vertex is called 'central'. The set of all central vertices is the center of the graph.

Biochemical Graphs

In modelling of metabolic pathways, graphs are often used to represent the large metabolic networks in various organisms. They have been extensively used to study the structural properties of metabolic regulation. For biochemical graphs, we can identify three types: the

molecule graph, the reaction graph, and the molecule-reaction graph. Below we will see how to construct each of them.

The *molecule graph* or metabolite graph, as it is called when talking about metabolic networks, is constructed by taking all molecules in a reaction system and connecting them if two molecules are in the same reaction.

The *reaction graph* is constructed in a similar way, but now we take all reactions as vertices and connect two of them if they contain the same molecule.

Finally, the combined *molecule-reaction graph* or metabolite-reaction graph is constructed by taking both molecules and reactions as vertices. If a molecule is in a reaction, we connect that molecule to the reaction. This specific graph can be made in an undirected way, as well as a directed way, where substrates are connected to reactions via outgoing edges towards reactions, while products are connected through incoming edges from reaction vertices.

In Figure 9, the three types of biochemical graphs are shown for the reaction system:

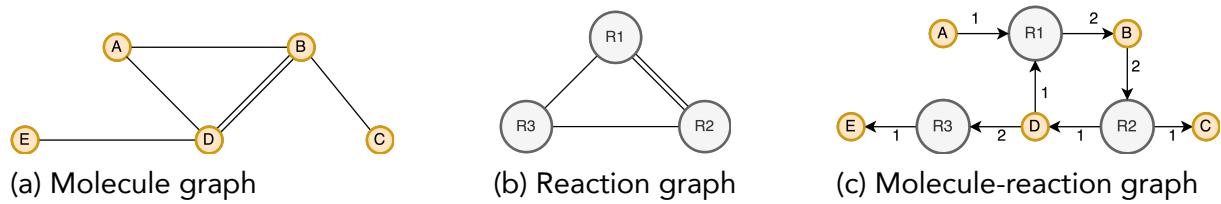


Figure 9 : Three types of graphs in biochemical systems.

Additionally, the combined molecule-reaction graph has a special property that the other two do not have. Notice that because of how the graph is constructed, we can never have two metabolites that are directly connected to each other, as well as no reactions that are connected to each other. A graph where you can color two “types” of notes, such that nodes of the same color have no edge between them is called *bipartite*.

Add example of bipartite graph

2.1.4 | Stoichiometry in Graphs

A common and useful property that is assigned to edges in molecule-reaction graphs is stoichiometry. In this way, the amount of molecules consumed and produced is included as information within the graph. Using this resulting directed graph with stoichiometric edge properties, we can construct the stoichiometric matrix. We can illustrate this using an example.

Example 2

We start with creating a matrix with the weights of all outgoing edges from each molecule towards the reaction, or equivalently, all incoming edges into each reaction.

$$A = \left(\begin{array}{c|ccccc} & A & B & C & D & E \\ \hline R_1 & 1 & 0 & 0 & 1 & 0 \\ R_2 & 0 & 2 & 0 & 0 & 0 \\ R_3 & 0 & 0 & 0 & 2 & 0 \end{array} \right) \quad (7)$$

Then, we will need a matrix with all incoming edges from each reaction towards each molecule, or equivalently, all outgoing edges from each reaction.

$$B = \left(\begin{array}{c|ccccc} & A & B & C & D & E \\ \hline R_1 & 0 & 2 & 0 & 0 & 0 \\ R_2 & 0 & 0 & 1 & 1 & 0 \\ R_3 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \quad (8)$$

Finally, we can compute the stoichiometry matrix (N) using the formula:

$$N = (B - A)^T \quad (9)$$

For this particular example, the stoichiometry matrix is given by:

$$N = \left(\begin{array}{c|ccc} & R_1 & R_2 & R_3 \\ \hline A & -1 & 0 & 0 \\ B & 2 & -2 & 0 \\ C & 0 & 1 & 0 \\ D & -1 & 1 & -2 \\ E & 0 & 0 & 1 \end{array} \right) \quad (10)$$

2.2 | Metabolic Networks

In systems biology, the analysis of complex systems of cells relies on the reconstruction of large metabolic networks. These metabolic networks contain many intermediate reactions, also referred to as *intermediary metabolism*. These reactions can be classified into two types of transformations: *catabolic* reactions break down complex substrates into simple metabolites, and *anabolic* reactions produce building blocks such as amino acids for the synthesis of proteins, nucleic acids for synthesis of DNA, RNA and energy intermediates such as ATP, and many others. All of these intermediate metabolic reactions take place through carrier molecules that together make up the metabolic network.

Metabolic networks are also hierarchical, where we can identify four levels. The first level contains the cellular inputs and outputs. In *in vitro* experiments, these inputs usually contain the medium and possible *perturbations*, and the outputs can be measured directly. In analysis of metabolic networks, we are often interested in the contribution of internal intermediates to the observed outputs.

The second level has already been described, and contains the two *sectors* of catabolism and anabolism. Often, another sector that can be identified is the property of *growth*.

The third level contains metabolic pathways, which can describe for example the uptake of glucose in a cell, its conversion to glucose-6-phosphate, and the further processing into the glycolysis pathway. Pathways are often important in pharmacology, where this specific level is often the target of drugs that stimulate or disturb an entire pathway.

The final level is the level of individual chemical reactions. Components at this level are often reconstructed to identify causes of diseases and their mechanisms. This is necessary in inborn metabolic errors, as they can be the result of enzymes being inactive, impacting one specific reaction in the human body. The reconstruction of these networks can be done using high-throughput sequencing and metabolomics data.

2.2.1 | Reconstruction of Metabolic Networks

To reconstruct these networks, various sources of information are often used. To obtain this network, a list of reactions is required. The most reliable source of information to obtain the list of reactions is direct biochemical information, from experiments where enzymes are isolated in an organism. For model organisms such as *E. coli* or yeast, this information can be readily available, but for humans, information is limited and hard to obtain. Alternatively, genome and epigenetic sequencing can reveal the transcribable proteins, and RNA sequencing data gives information on the actual genes being transcribed at a specific moment in time. Using perturbation experiments, the transcription reaction to molecules can be identified. More modern sequencing techniques, such as ChIP-seq can also give information on the DNA-protein interactions.

Many online databases exist for all of the data types mentioned, and for some organisms, large reconstructions of metabolic networks exist. Examples include:

1. [EMBL-EBI \(genome data\)](#)
2. [UniProt \(protein data\)](#)
3. [Enzyme commission numbers](#)

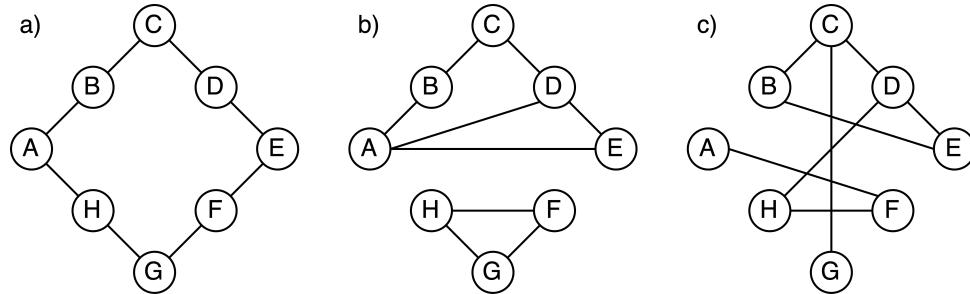
Additionally, known physiological constraints, such as localization constraints, or the biological knowledge that an organism or cell can produce a specific enzyme can help in identifying missing components. Finally, we can fill up missing components by using computer models. As our network needs to be able to simulate, we can fill in the gaps by fixing mass balances, or adding molecules necessary to produce specific products. These unvalidated reactions are called *inferred* reactions. The computational tools used to infer these reactions are beyond the scope of this course and will be discussed in the Systems Medicine master course.

2.2.2 | The Metabolic Network of Yeast

Up to now, these networks have been particularly useful in bioengineering contexts, where DNA modifications are introduced in model organisms to enable the production of molecules or drugs that are otherwise hard to synthesize. Additionally, the identification of the origins of metabolic diseases is an important application of these networks.

2.3 | Exercises

1. For the following undirected graphs, write down the adjacency matrix



2. For graph c) in the previous exercise:

- a) compute the eccentricity of each vertex.
- b) What is the radius of this graph?
- c) What is the diameter?
- d) Which of the vertices are part of the centre?

3. Take graph c) from exercise 1:

- a) How many walks exist from B to D of length 2
- b) And how many of length 3?
- c) And how many of length 4?

4. Given the adjacency matrix of graph c) of exercise 1 as C , compute the following matrix products:

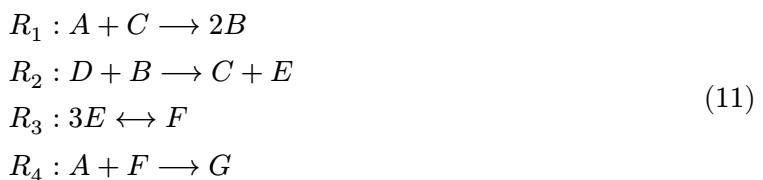
Hint: Use MATLAB or Python + numpy to speed up the computation.

- a) $C \cdot C$
- b) $C \cdot C \cdot C$
- c) $C \cdot C \cdot C \cdot C$

Compare the entries at point B,D to your answers in question 3. What do you observe?

5. Give the molecule, reaction and molecule-reaction graphs for the following reaction system.

Add stoichiometric coefficients to the edge weights of the molecule-reaction graph.



6. Assemble the Stoichiometric Matrix of the system in question 5.

"The only thing predictable about life is its unpredictability."

— Remi (Ratatouille)

At the end of the previous chapter, we have seen that we can express systems of chemical reactions as various types of graphs. We closed with the molecule-reaction graph, where the stoichiometric coefficients represent the edge weights. We also introduced the stoichiometry matrix, or N , which could be derived from the substrate-reaction graph. In this chapter, we will use this graph, combined with other concepts to discuss dynamic behavior of these reaction systems.

We will also dive into systems that cannot be simply written as linear systems of equations, and we will delve into some examples of these systems. Furthermore, some recognizable elements of these dynamic systems are discussed separately. This is especially important when describing a model of two competitive binders, enzyme kinetics, or when we are modeling systems that automatically restore to their original state with the help of feedback loops. This will be heavily illustrated with examples, which also show how these feedback loops can sometimes explain mechanisms of pathology.

Additionally, some systems become stable if no external input is given, which is called the steady-state. We will discuss this steady-state and so-called model setpoints that specifically determine the steady-state of a system. These inputs that can be used to push a model out of its steady state are called perturbations. We will show examples of how these can be modeled and what kind of perturbations are often used.

Finally, we will discuss how these models can be combined with experimental measurements to validate or create new hypotheses about system behavior. This will be accompanied by a discussion of pharmacological modelling, and how these dynamic pharmacokinetic models play a large role in describing how drugs are distributed throughout the human body.

3.1 | Biochemical Systems

As discussed in the Introduction to this chapter, we will go more in-depth into the chemical reactions discussed in the previous chapter. We will start with an example.

Example 3

Observe the directed molecule reaction graph shown in Figure 10. This graph represents the reaction: $2X \rightarrow Y$

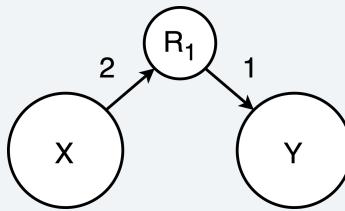


Figure 10 : The directed molecule reaction graph of a potential dimerization of monomer X into dimer Y . The edge weights represent the stoichiometric coefficients for the reaction.

This can for example represent a simple dimerization of two monomers X into a dimer Y . Using the graph, we can construct our substrate matrix, which is given by the weight of all the outgoing edges for each molecule and reaction:

$$A = \begin{pmatrix} X & Y \\ R_1 & 2 & 0 \end{pmatrix} \quad (12)$$

And we can construct the product matrix, where we take all the weights of incoming edges for each molecule and reaction:

$$B = \begin{pmatrix} X & Y \\ R_1 & 0 & 1 \end{pmatrix} \quad (13)$$

Using these two matrices, we can construct the stoichiometry matrix.

$$N = (B - A)^T = \begin{pmatrix} R_1 \\ X & -2 \\ Y & 1 \end{pmatrix} \quad (14)$$

This procedure can give us static information about the reaction system we are studying, such as whether a molecule is produced or consumed overall, and which molecules may be most critical for the system. However, if we want to observe how this reaction occurs over time, we do not have enough information yet. For this, we are going to look at *reaction rates* in the next part of this section.

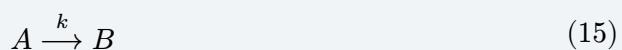
3.1.1 | The Law of Mass-Action

To be able to describe concentrations of molecules in a reaction system over time, we need to know at what rate the reaction occurs, besides the stoichiometry of the system. This reaction rate is defined as the amount of molecules that is converted per unit of time. The reaction rate is a row-vector with an entry for each reaction and labelled v .

The reaction rate for each reaction is given by the law of mass-action, which states that the rate of any reaction is directly proportional to the concentrations of our substrates. This means that for the reaction rate, we multiply all substrates together, and multiply this with the basal reaction rate k .

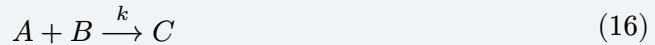
Example 4

1. For the reaction



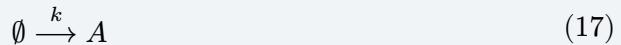
We get the reaction rate $v = k[A]$

2. For the reaction



We get the reaction rate $v = k[A][B]$

3. For the reaction

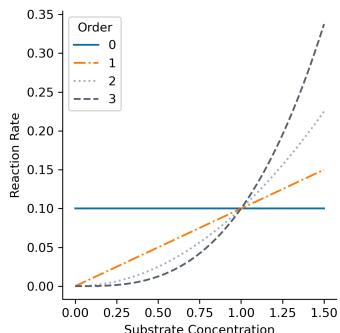


We get the reaction rate $v = k$, as we have no substrates.

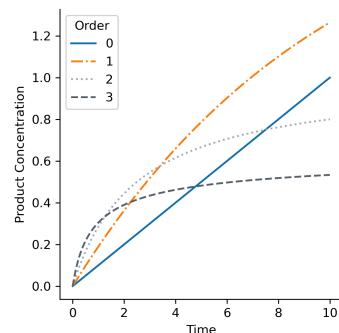
When looking at this example, we can use the same rules to identify the reaction rates from our dimerization example from Example 3. In that example, the reaction rate is given by

$$v = k[X][X] = k[X]^2 \quad (18)$$

The *order* of a reaction is the highest exponent we can find in the reaction rate. For example, the first two reactions from Example 4 are both *first order* reactions, and the last reaction from that example is a *zeroth order* reaction. Figure 11a shows the reaction rates for different reaction orders, with a rate constant of $k = 0.1$. We see that lower order reactions typically have higher reaction rates for substrate concentrations below 1, while higher order reactions have higher reaction rates above substrate concentrations of 1. In Figure 11b you can subsequently see the production of B over time for the different reaction orders.



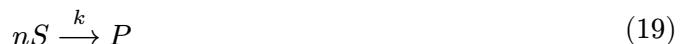
(a)



(b)

Figure 11 : Simulations of reaction orders 0 to 3. **(a)** The reaction rate according to mass-action kinetics for different reaction orders for a rate constant of $k = 0.1$. **(b)** Simulation over time of the production of B in the reaction $nA \xrightarrow{k} B$ for various reaction orders.

If we take things to a more general level, the law of mass action dictates that for a reaction in the form of:



the reaction rate of formation of the product P is given by:

$$v = k[S]^n \quad (20)$$

For a reaction with x substrates, the reaction rate of the product is given by:

$$v = k([S]_1^{n_1} \cdot [S]_2^{n_2} \cdots [S]_x^{n_x}) \quad (21)$$

In this way, we can use the law of mass action in combination with the reaction scheme to construct the reaction rates.

3.1.2 | The Rate Equation

Using the stoichiometry matrix and the rates computed using the law of mass-action, we can create the *rate equation*. This is a differential equation describing the change of our concentrations over time.

The rate equation is given by

$$\frac{du}{dt} = N \cdot v \quad (22)$$

This is a differential equation. Filling in the stoichiometry matrix and the rate, we get

$$\begin{aligned} \frac{d[X]}{dt} &= -2 \cdot k[X]^2 \\ \frac{d[Y]}{dt} &= k[X]^2 \end{aligned} \quad (23)$$

We will now take a brief intermezzo into differential equations.

Differential Equations

Instead of directly describing molecular modelling processes as functions of time, it is often more intuitive to use *ordinary differential equations* to describe the system's behavior over time. Differential equations have been introduced to you in your calculus course, and you may be familiar with solving simple examples. However, we will not be concerned with solving these differential equations analytically. Instead, we will analyze them directly and describe their components.

An intuitive way of looking at a differential equation for a molecule X can be:

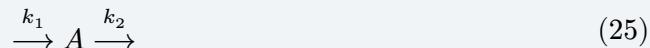
$$\underbrace{\frac{dX}{dt}}_{\text{rate of change of } X \text{ at time } t} = \text{Production}(X) - \text{Consumption}(X) \quad (24)$$

Often, we can derive these production and consumption terms from mass-action kinetics, but as modelers, we also have the freedom of taking other assumptions about these terms. Examples of these will be discussed in the next section. For now, let's further explore some examples of mass-action kinetic models and identify the production and consumption terms in each model.

In our first example, let's discuss a system with two reactions and one molecule. This type of reaction is a very simple basis for a biochemical model, where we have a constant input, and a concentration-dependent consumption term. In future sections, you will be able to recognize the result of this example in parts of models.

Example 5

Let's consider the reaction



We only have one molecule, but we have two reactions. Therefore, the substrate matrix looks like:

$$A = \begin{pmatrix} & A \\ R_1 & 0 \\ R_2 & 1 \end{pmatrix} \quad (26)$$

And our product matrix looks like:

$$B = \begin{pmatrix} & A \\ R_1 & 1 \\ R_2 & 0 \end{pmatrix} \quad (27)$$

We can construct our stoichiometry matrix as:

$$N = (B - A)^T = \begin{pmatrix} & R_1 & R_2 \\ A & 1 & -1 \end{pmatrix} \quad (28)$$

We now only need our rates, the rate for R_1 is equal to k_1 , as we have no substrates, but the rate for R_2 is equal to $k_2[A]$, as we have A as its substrate. The rate vector then looks like:

$$v = \begin{pmatrix} R_1 & k_1 \\ R_2 & k_2[A] \end{pmatrix} \quad (29)$$

Our final differential equation then looks like:

$$\frac{d[A]}{dt} = N \cdot v = (1 \ -1) \cdot \begin{pmatrix} k_1 \\ k_2[A] \end{pmatrix} = k_1 - k_2[A] \quad (30)$$

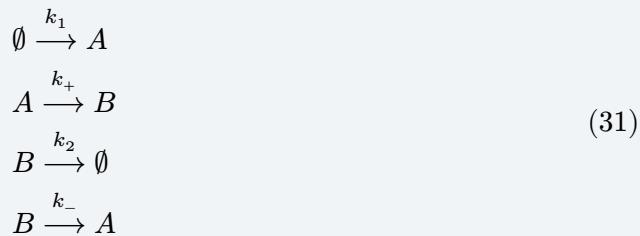
In the example above we have shown two reactions with one molecule. We can also create an example of two molecules of which one is produced and the other one is consumed in accordance with both reactions of the example above. However, these molecules can also be converted into one another. Let's see how mass-action kinetics can help us define a differential equation for this system.

Example 6



Figure 12 : A reaction system with two molecules and one two-sided reaction.

Consider the system that is shown in Figure 12. We can write the individual reactions as:



We now have four reactions and two molecules. Therefore, the substrate matrix looks like:

$$A = \begin{pmatrix} & A & B \\ R_1 & 0 & 0 \\ R_2 & 1 & 0 \\ R_3 & 0 & 1 \\ R_4 & 0 & 1 \end{pmatrix} \tag{32}$$

And our product matrix looks like:

$$B = \begin{pmatrix} & A & B \\ R_1 & 1 & 0 \\ R_2 & 0 & 1 \\ R_3 & 0 & 0 \\ R_4 & 1 & 0 \end{pmatrix} \tag{33}$$

We can construct our stoichiometry matrix as:

$$N = (B - A)^T = \begin{pmatrix} & R_1 & R_2 & R_3 & R_4 \\ A & 1 & -1 & 0 & 1 \\ B & 0 & 1 & -1 & -1 \end{pmatrix} \tag{34}$$

Now what about the rates. Our first and second reactions are similar to the previous example, and we can also easily use the mass-action theory to get the other rates, to obtain the rate vector:

$$v = \begin{pmatrix} R_1 & k_1 \\ R_2 & k_+[A] \\ R_3 & k_2[B] \\ R_4 & k_-[B] \end{pmatrix} \tag{35}$$

Combining both using the rate equation will result in two differential equations. We get:

$$\begin{aligned}
 \frac{d[A]}{dt} &= k_1 - k_+[A] + k_-[B] \\
 \frac{d[B]}{dt} &= k_+[A] - (k_2 + k_-)[B]
 \end{aligned} \tag{36}$$

In this example, we can see that A contained two production terms and one consumption term, while B contained two consumption terms and only one production term. In the next example, we will get a little closer to a metabolic process. When looking at Figure 13 you

may be able to imagine A as a molecule outside the cell, and C being the same molecule, but inside the cytosol. B and D could represent ATP and ADP respectively, and we have a very simple model of active transport. The reaction from D to B is a large oversimplification in the case of ADP and ATP, as the regeneration involves multiple steps.

Example 7

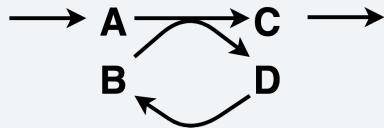


Figure 13 : A cyclic reaction system, resembling a simple metabolic cycle.

As a final example, observe the following system of reactions:



This system is also shown in Figure 13. For this system, we can determine the differential equations in the same way as for the previous two systems:

$$\begin{aligned} \frac{d[A]}{dt} &= k_1 - k_2[A][B] \\ \frac{d[B]}{dt} &= k_3[D] - k_2[A][B] \\ \frac{d[C]}{dt} &= k_2[A][B] - k_4[C] \\ \frac{d[D]}{dt} &= k_2[A][B] - k_3[D] \end{aligned} \tag{38}$$

Identifying Components of Biochemical ODE Systems

In the previous part, you may have seen repeated elements in these systems of differential equations. Elements that couple two of the described molecules, or more formally called *state variables*, are called *coupling terms*. Because of these, our systems become complicated to solve, as each state variable could be dependent on many others. Besides the identification of coupling terms, it is important to be able to distinguish production from consumption terms. Observe the system from example 3.3 and shown in Figure 13. When looking at the differential equations, we can see that all of our state variables are coupled by the term $k_2[A][B]$. Furthermore, $[B]$ and $[D]$ are coupled by the term $k_3[D]$. In biochemical systems, these coupling terms contain consumptive and productive counterparts.

Other terms in the system are open-ended, and are either production or consumption terms, such as the constant production of $[A]$, governed by term k_1 . In this way, each term in the ODE system can be explained. This explainability is important, as a direct link to the

underlying chemical or biological process can be made. Using this knowledge, we can perform targeted hypothesis testing by simulating these systems under different conditions. This simulation will be discussed later in these lecture notes. First, we will take a look at a different type of system: the signalling network.

3.2 | Signaling Systems

In the previous section, we have observed systems of molecules that are produced and consumed according to mass-action kinetics. In this section, we will see that we can use the mass-action formalism to model stimulatory and suppressive signals. However, we will also show some assumptions that can be made to simplify the model. In this section, we will talk about ways to introduce suppressive or stimulatory signals into a model, and illustrate some differences with earlier biochemical models. We will then zoom out of individual signals and inspect models as a whole. In particular, we'll discuss feedback loops that arise as a consequence of included signalling pathways and end with an example model containing several components we have discussed in this and the previous section.

3.2.1 | Linear Stimulation

The first kind of signalling that we can model is linear stimulation. Observe the reaction



For this reaction, we can see that there is no net consumption of X , and we can use mass action kinetics to describe the conversion of A and B , which is mediated by X in this case:

$$\begin{aligned} \frac{d[A]}{dt} &= -k[A][X] \\ \frac{d[B]}{dt} &= k[A][X] \\ \frac{d[X]}{dt} &= 0 \end{aligned} \quad (40)$$

We can see from the equations that in presence of X , molecule A is converted into molecule B , which is a form of positive interaction. In this form, we assume that the more of X we have, the quicker A is converted into B , without a limit. However, when X is not present, we see no conversion. We can also see this in Figure 14, which shows that with increasing value of our stimulatory agent $[X]$, the rate of conversion of A into B increases, but without X , we see no conversion happening.

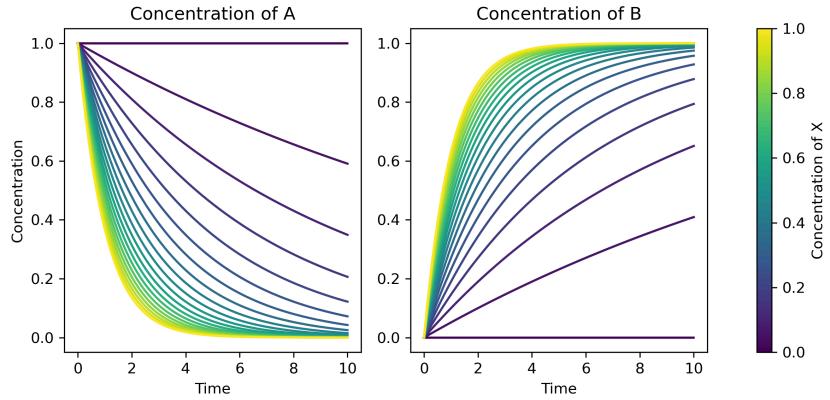


Figure 14 : Simulation of a linear stimulation model where A is converted into B using a stimulatory agent X . The colors indicate the concentrations of this stimulatory agent.

A method to extend this model is to add the possibility for A to spontaneously convert into B without the interaction with X . We can do this by adding the reaction:



Combining both into a system of differential equations will result in:

$$\begin{aligned} \frac{d[A]}{dt} &= -(k[X] + k_{sp})[A] \\ \frac{d[B]}{dt} &= (k[X] + k_{sp})[A] \\ \frac{d[X]}{dt} &= 0 \end{aligned} \quad (42)$$

In this model, we can see that the rate of conversion from A to B is subject to a basal rate k_{sp} and increases linearly with the concentration of X .

3.2.2 | Linear Suppression

Besides stimulation, we can also model suppression in a similar way. However, we will need some more tools to do so. A way to model suppression is that we have a reaction where an active form of molecule A is converted into molecule B :



Additionally, we introduce a suppressive agent Y , which blocks this conversion. This can be modelled by a reaction, mediated by Y , that converts our active form of A into an inactive form:



We now still miss one key component in this model. We want this suppression to be reversible, meaning that as soon as Y disappears, we need the conversion of A into B to occur again. Therefore, we can add a reaction where A_{inact} converts into A_{act} :



Combining these three reactions into a system of differential equations can once again be done using mass-action kinetics:

$$\begin{aligned}\frac{d[A_{\text{act}}]}{dt} &= -(k + k_i[Y])[A_{\text{act}}] + k_a[A_{\text{inact}}] \\ \frac{d[A_{\text{inact}}]}{dt} &= k_i[Y][A_{\text{act}}] - k_a[A_{\text{inact}}] \\ \frac{d[B]}{dt} &= k[A_{\text{act}}] \\ \frac{d[Y]}{dt} &= 0\end{aligned}\tag{46}$$

Observe that we have added an additional equation that specifies that the concentration of A equals the sum of active and inactive substrate. A simulation of this system for different values of our suppressant Y can be seen in Figure 15. If we increase the level of suppressant, we see that the proportion of A being converted into B decreases.

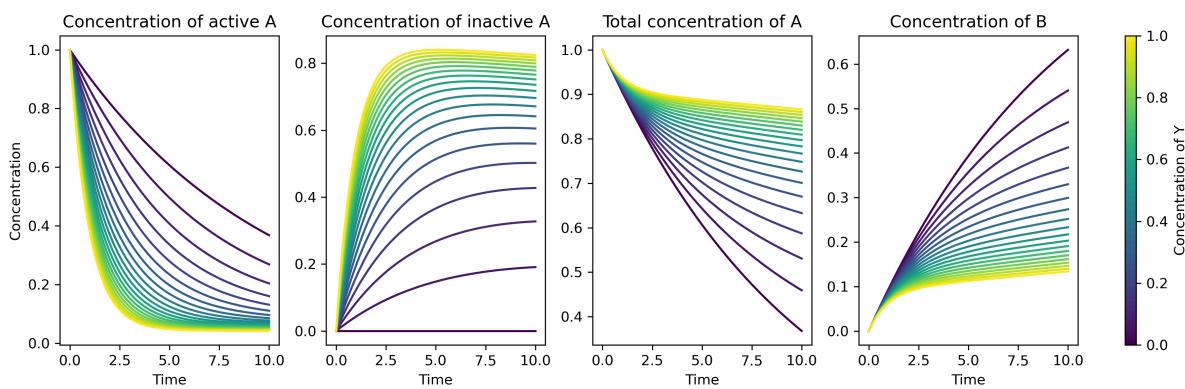


Figure 15 : Simulation of a linear suppression model with suppressant Y . The colors indicate the concentration of this suppressant.

3.2.3 | Enzyme Kinetics

A special case of stimulation or suppression can be seen in enzyme catalysis. In your biochemistry course, you will have seen Michaelis-Menten kinetics. For completeness, the derivation of this type of kinetic model is given here, but the most important consideration is that you realize the assumption it makes and what the kinetic equation looks like, so you can recognize it in future models. We will first start with the derivation of the kinetic law using a stimulatory enzyme.

An enzymatic reaction can be summarized as:



In this reaction, an enzyme E and a substrate S form a complex C , which can also fall back apart. However, if the catalysis succeeds, the enzyme is released and a product P is formed. Using mass-action kinetics from the previous sections, we can create a differential equation model for this system:

$$\begin{aligned}
\frac{d[S]}{dt} &= k_{-1}[C] - k_1[E][S] \\
\frac{d[E]}{dt} &= (k_{-1} + k_2)[C] - k_1[E][S] \\
\frac{d[C]}{dt} &= k_1[E][S] - (k_{-1} + k_2)[C] \\
\frac{d[P]}{dt} &= k_2[C]
\end{aligned} \tag{48}$$

The main assumption for Michaelis-Menten kinetics, is that the total conversion from substrate into product is mainly determined by the formation of the complex C . The consequence of this assumption, is that the amount of complex in the system rapidly reaches an equilibrium, which results mathematically into the relationship:

$$\frac{d[C]}{dt} = 0 \tag{49}$$

Using this assumption, we can write:

$$k_1[E][S] = k_{-1} + k_2[C] \tag{50}$$

Furthermore, the second assumption is that the total amount of enzyme in the system doesn't change. The total amount of enzyme can then be formulated as the sum of the amount of free enzyme $[E]$ and the amount of complex $[C]$. We can call this concentration E_0 :

$$E_0 = [E] + [C] \tag{51}$$

Replacing $[E]$ in Equation 50 with $E_0 - [C]$ gives us:

$$k_1 E_0 [S] = (k_{-1} + k_2 + k_1 [S]) [C] \tag{52}$$

From which we can derive the following formula for $[C]$:

$$[C] = \frac{k_1 E_0 [S]}{k_{-1} + k_2 + k_1 [S]} \tag{53}$$

Dividing both numerator and denominator by k_1 and defining $K_M = \frac{k_{-1} + k_2}{k_1}$ we get:

$$[C] = \frac{E_0 [S]}{K_M + [S]} \tag{54}$$

Filling this in into the function for the product formation, and defining

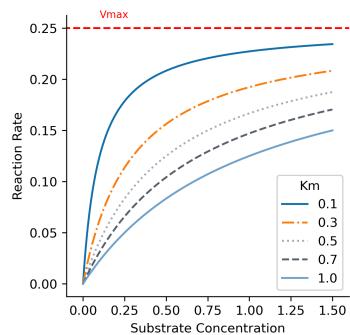
$$V_{\max} = k_2 E_0 \tag{55}$$

we get:

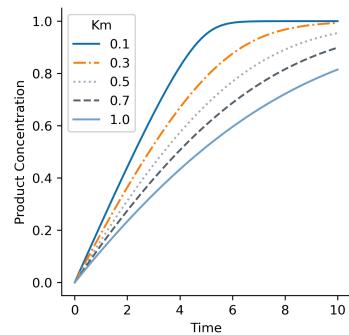
$$\frac{d[P]}{dt} = V_{\max} \frac{[S]}{K_M + [S]} \tag{56}$$

The rate in Equation 56 is what we define as Michaelis-Menten kinetics. The behavior of this type of kinetics can be observed in Figure 16a, which shows the reaction rate as a function of the substrate concentration for different values of K_M . You can see that for increasing substrate concentration, the reaction rate approaches V_{\max} . In Figure 16b you can see the

product concentration over time from a simulation of the Michaelis-Menten conversion. This conversion is also compared to the linear mass-action conversion in this figure. The effect of K_M is also clearly visible in both figures.



(a)



(b)

Figure 16 : Simulation of Michaelis-Menten kinetics for various values of K_M . (a) The reaction rate according to Michaelis-Menten kinetics as a function of the substrate concentration, compared to the value of V_{max} . (b) Simulation of product formation according to Michaelis-Menten kinetics.

3.2.4 | Advanced Enzyme Kinetics: Reversible Inhibition

Sometimes, an enzymatic reaction can occur in presence of an inhibitor I . As discussed in earlier courses, reversible inhibition can take three forms. We have competitive inhibition, non-competitive inhibition and uncompetitive inhibition (see also Figure 17). These three forms also have distinct kinetic rates that we can derive. Their derivations are beyond the scope of these lecture notes, but their resulting equations can be readily explained from the mechanisms of each type of inhibition.

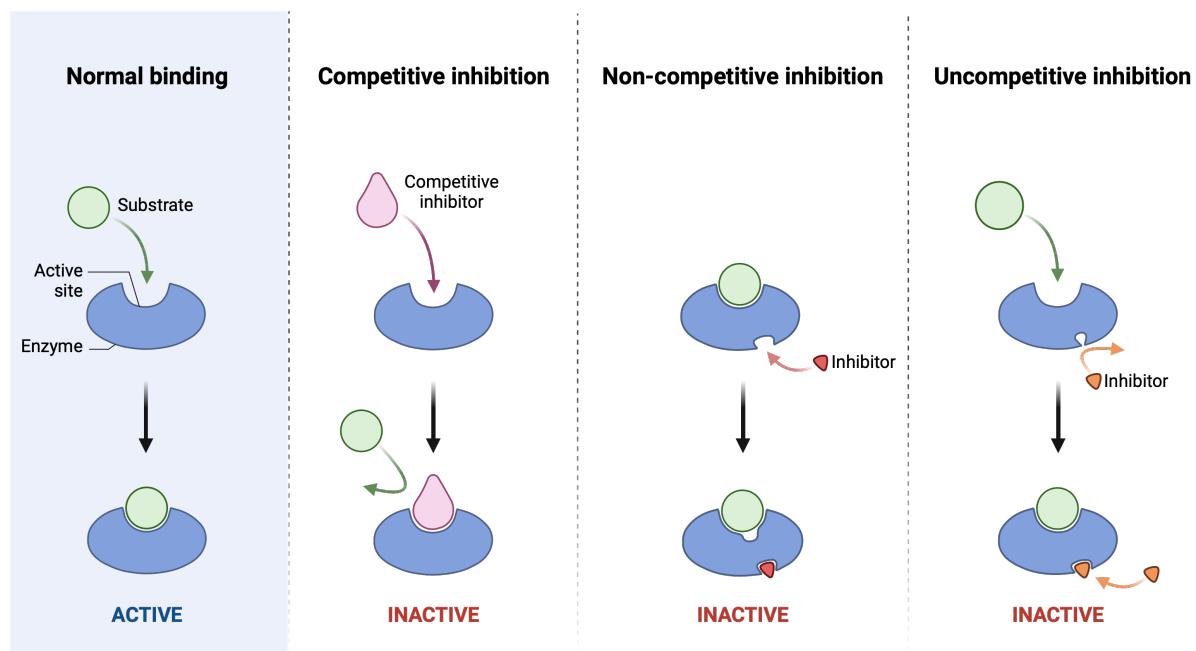
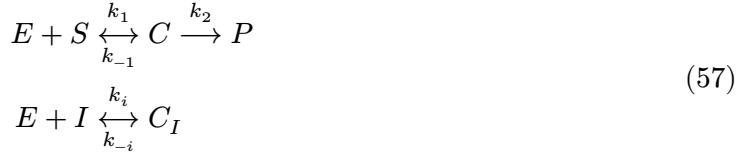


Figure 17 : Forms of reversible inhibition. With competitive inhibition, the inhibitor binds to the active site of the substrate, preventing the formation of an active complex. In non-competitive inhibition, the inhibitor binds to an allosteric site, reducing or blocking the catalytic activity of the enzyme. In uncompetitive inhibition, the inhibitor can only bind to an allosteric site of the enzyme-substrate complex.

Competitive Inhibition

Competitive inhibition can be represented in chemical reactions as



Where we have a normal complex C and a substrate-inhibitor complex C_I . The rate law for competitive inhibition with an inhibitor I is given by:

$$v = \frac{V_{\max}[S]}{K_M^{\text{app}} + [S]} \quad (58)$$

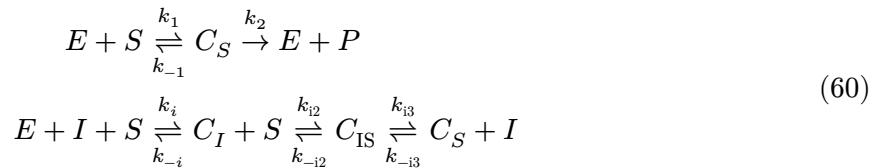
With

$$K_M^{\text{app}} = K_M \left(1 + \frac{[I]}{K_i} \right) \quad (59)$$

The result of this competitive inhibition on the reaction rates can be seen from the equation. As we increase the inhibitor concentration $[I]$, we see that apparent Michaelis-Menten constant (K_M^{app}) increases, meaning that we will need more substrate to reach our V_{\max} , (also see Figure 16) but we will still be able to reach this maximum reaction rate. This effect can be explained as the presence of this inhibitor decreases the enzyme affinity for the substrate. Many drugs act as competitive inhibitors, as they are designed to resemble the substrate and therefore bind to or block the active site of specific enzymes.

Non-competitive Inhibition

The second type of inhibition is non-competitive inhibition, which can be formulated as



This reaction system shows that our inhibitor can bind and unbind from every step of the catalytic process. For a non-competitive inhibitor, we will see that instead of increasing K_M , we decrease V_{\max} with the rate law:

$$v = \frac{V_{\max}^{\text{app}}[S]}{K_M + [S]} \quad (61)$$

With

$$V_{\max}^{\text{app}} = \frac{V_{\max}}{1 + \frac{[I]}{K_i}} \quad (62)$$

As the binding of a non-competitive inhibitor reduces the activity of the enzyme without changing the enzyme's affinity for the substrate, the V_{\max} is decreased. Examples of non-competitive inhibition include the binding of Glucose-6-Phosphate to hexokinase in the brain, slowing the rate of cerebral glucose uptake.

Uncompetitive Inhibition

The final type of inhibition is called uncompetitive inhibition, which changes both V_{\max} and K_M . We can formulate this inhibition as:



We can see that the inhibitor can only bind to the enzyme-substrate complex. This type of inhibition is very rare, but does occur. The rate law for this type of inhibition is given by:

$$v = \frac{V_{\max}^{app}[S]}{K_M^{app} + [S]} \quad (64)$$

With

$$V_{\max}^{app} = \frac{V_{\max}}{1 + \frac{[I]}{K_i}} \quad (65)$$

and

$$K_M^{app} = \frac{K_M}{1 + \frac{[I]}{K_i}} \quad (66)$$

Contrarily to competitive inhibition, the K_M decreases with increasing inhibitor concentration, while V_{\max} decreases, similarly to non-competitive inhibition. One of the main features of this type of inhibition is that its effect is largest at high substrate concentrations.

3.2.5 | Cooperative Binding: Hill Kinetics

The production rate of the product P described by Michaelis–Menten kinetics, is a hyperbolic function of the substrate concentration $[S]$. However, in many cases, measuring the reaction rate as function of the substrate concentration leads to different behavior. One reason may be cooperative binding of the substrate to the enzyme. Cooperative binding means that the binding of one substrate molecule has influence on the binding of subsequent substrate molecules to the enzyme (possibly, with several active sites). The cooperativity is positive if the binding of a first substrate molecule increases the affinity of other active sites of the enzyme for substrate molecules, and negative if this binding decreases the affinity of other active sites. The effective rate of a reaction with such a cooperative effects is often described by an equation similar to the Michaelis-Menten rate equation:

$$\frac{d[P]}{dt} = V_{\max} \frac{[S]^n}{(K_M)^n + [S]^n} \quad (67)$$

The difference is that we have introduced an additional term n , in the exponent of every term in the fraction. This type of kinetics is called *Hill kinetics*, and Equation 67 is called the *Hill equation*.

For $n = 1$, this equation is equal to the Michaelis-Menten kinetics. This value n is linked to the cooperativity of the reaction. A value of n in between 0 and 1 indicates a negative cooperativity, while a value for $n > 1$ indicates positive cooperativity.

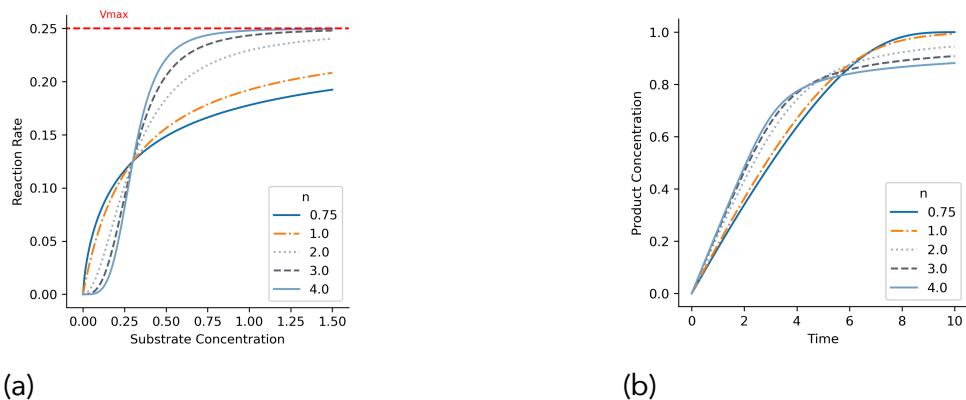


Figure 18 : Simulation of Hill kinetics for various values of n . **(a)** The reaction rate according to Hill kinetics as a function of the substrate concentration, compared to the value of V_{max} . **(b)** Simulation of product formation according to Hill kinetics.

3.2.6 | Feedback Loops

Combining these regulatory mechanisms that are present in models of biological systems often leads to feedback loops appearing. A feedback loop is a specific regulatory pattern that dictates how components of a system interact over time. They are an important component of many biological systems and are required for a system to return to its original state after an external influence, or cause repetitive behavior to occur.

We can identify positive and negative feedback loops. A positive feedback loop occurs when the endpoint of a series of reactions promotes the starting point of this same series, causing the complete set of reactions to become increasingly active over time. When the endpoint blocks the starting point of this series, we call this negative feedback, which is necessary for a system to return to its original state.

Combinations of positive and negative dictate our bodily processes, and a disturbance in the balance of these loops can lead to diseases. Examples include stress-related diseases or diabetes. When analyzing a biological system, it often helps to outline the positive and negative feedback loops in the system to get an understanding of the interacting processes.

illustration?

3.2.7 | Modelling Example: The Cell Cycle

In this part, we'll be looking at an example model, explaining the components using the modelling tools we have explored up to this point. The following model is a heavily simplified model of the cell cycle published in 1991 by Goldbeter [7]. In this model, cyclin (C) is modelled to induce the production of a cyclin kinase (M), which in turn activates the production of a cyclin protease (X). This protease then stimulates the degradation of the original cyclin. An illustration of the model and its interactions is given in Figure 19.

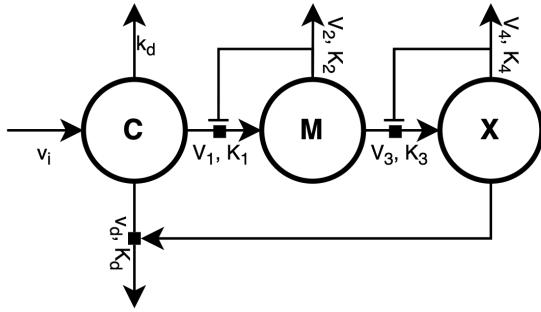


Figure 19 : Illustration of the components and their interactions in the Goldbeter model cite{Goldbeter1991} of the cell cycle. The figure shows interactions between cyclin (C), cyclin kinase (M) and cyclin protease (X). Arrows indicate positive interactions and bars indicate suppressions. Parameter names are given for all interactions.

The model is mathematically formulated as:

$$\begin{aligned}\frac{dC}{dt} &= v_i - v_d X \frac{C}{K_d + C} - k_d C \\ \frac{dM}{dt} &= V_1 \frac{C}{K_c + C} \frac{1 - M}{K_1 + 1 - M} - V_2 \frac{M}{K_2 + M} \\ \frac{dX}{dt} &= V_3 M \frac{1 - X}{K_3 + 1 - X} - V_4 \frac{X}{K_4 + X}\end{aligned}\quad (68)$$

We can see that the model contains six examples of Michaelis-Menten kinetics. The equations for cyclin kinase and cyclin protease (M and X) are structurally similar. We will first look at cyclin (C). The base of this equation is the constant production and removal term, according to mass action kinetics, and based upon the chemical formulation:



We then see that the additional term represents a catalytic consumption of cyclin by cyclin protease, yielding an unmodelled product. As the enzyme concentration (X) varies throughout the simulation, the previously constant term V_{\max} is replaced by $v_d X$, using the original definition of V_{\max} (see Equation 55).

Cyclin kinase (M) production is then stimulated by cyclin using a Michaelis-Menten factor in the first term of the second reaction. The term after this factor represents Michaelis-Menten kinetics of a cyclin kinase progenitor, where the total concentration of cyclin kinase and its progenitor is normalized to the constant value of 1. We can then describe the progenitor concentration as $1 - M$. The second term describes the consumption of cyclin kinase by an unmodelled enzyme with constant concentration using Michaelis-Menten kinetics.

The equations of cyclin protease (X) have the same structure as cyclin kinase, where its production is stimulated by cyclin kinase, and is consumed using Michaelis-Menten kinetics, driven by an unmodeled enzyme with constant concentration.

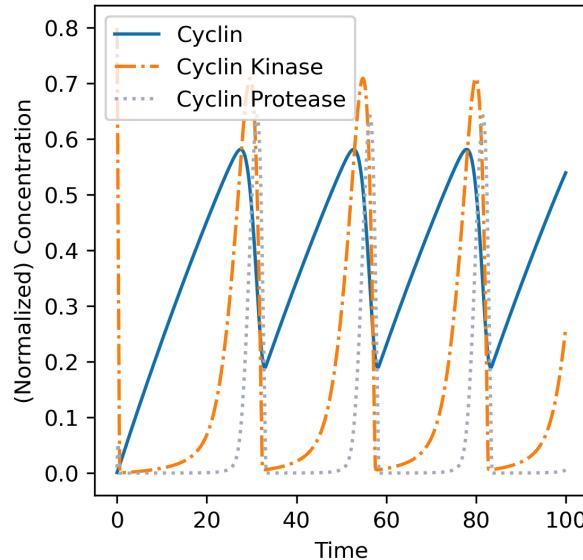


Figure 20 : Simulation of the Goldbeter model. We can clearly see the oscillatory behavior that the model produces.

When simulating this model, we find that it shows oscillations in all three state variables (see Figure 20). This behavior is a result of the negative feedback loops present in the system. The system is kept running through the parameters v_i and k_d , which provide a constant production and consumption of C . Then C stimulates M , which is suppressed by M . Then, M stimulates X , which is suppressed by X . Finally, X suppresses C by stimulating an additional consumption term, allowing for a restart of the cycle.

3.3 | Steady-States, Setpoints, and Perturbations

Besides the simulation, we may often be interested in other model properties that we can calculate. An important property of dynamic models is the steady state. When we have a dynamic model defined by:

$$\frac{dx}{dt} = f(x) \quad (70)$$

The steady state is given by $f(x) = 0$, which means that our state variable does not change over time anymore. We can, for example, compute the steady-state for the model from Equation 46.

Example 8

The system represents a linear suppression model with the equations:

$$\begin{aligned} \frac{d[A_{\text{act}}]}{dt} &= -(k + k_i[Y])[A_{\text{act}}] + k_a[A_{\text{inact}}] \\ \frac{d[A_{\text{inact}}]}{dt} &= k_i[Y][A_{\text{act}}] - k_a[A_{\text{inact}}] \\ \frac{d[B]}{dt} &= k[A_{\text{act}}] \\ \frac{d[Y]}{dt} &= 0 \end{aligned} \quad (71)$$

From our first analysis, we see that Y is always in steady state, as its derivative is set to zero. To compute the steady-state for the entire system, we find that B is only in steady-state whenever $[A_{act}]$ is zero. Using this fact, we can express the steady state of A_{act} as

$$\frac{d[A_{act}]}{dt} = \underbrace{-(k + k_i[Y])[A_{act}]}_{=0, \text{ since } [A_{act}]=0} + k_a[A_{act}] = k_a[A_{inact}] = 0 \quad (72)$$

From this we conclude that in steady-state $[A_{inact}] = 0$ as well. Filling this in, we also see that no further conditions are necessary to make sure $[A_{inact}]$ is also in steady state.

Systems can also have *multiple* steady-states. A system with two steady-states is also called *bistable*. In biological systems, this property of bistability is important in understanding specific disease mechanisms.

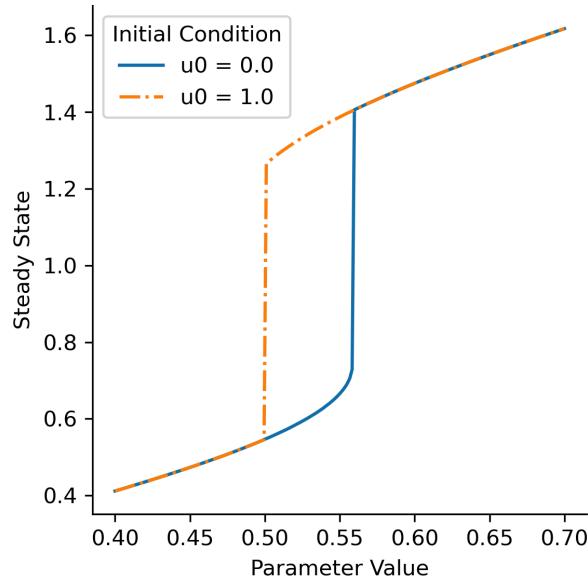


Figure 21 : The steady-states in a bistable system.

Bistability can be caused by a positive feedback loop where one regulatory step is very sensitive. An example is the following equation.

$$\frac{dy}{dt} = p + \frac{y^5}{1+y^5} \quad (73)$$

In this system, we can also recognize a Hill-kinetics rate equation. The steady states of this system depending on parameter value p are shown in Figure 21. The point where the steady-state splits based on initial condition is called a *bifurcation point*, as over there, the steady-state splits into two possibilities based on the initial condition.

A different property of a system is called a *setpoint*. In control systems, a setpoint is the target value of a specific variable in the entire dynamic system. Setpoints are often used when modeling biological systems that typically return to a specific steady state, such as glucose, which typically has a fasting value of around 5 mmol/L. In modeling, we then use a setpoint to force the system into steady state as soon as this is the case. An example can be seen in a simple glucose model.

Example 9

The Bergman glucose minimal model [8] can be formulated as:

$$\begin{aligned}\frac{dG}{dt} &= -k_1 GX - k_3(G - G_b) + Ra(t) \\ \frac{dX}{dt} &= -X + k_2(I(t) - I_b)\end{aligned}\tag{74}$$

This model describes the basic regulation of glucose by insulin. The components of insulin ($I(t)$) and glucose input ($Ra(t)$) are described as external inputs to the model, so they must be known to describe the whole system.

When investigating this system, we see that the following conditions need to be met for the system to be in steady state:

- $X = 0$
- $I(t) = I_b$
- $G = G_b$
- $Ra(t) = 0$

The model contains two setpoints; we have I_b , which is the setpoint of the insulin concentration, and G_b , which is the setpoint of the glucose concentration.

In some model applications, we also want to provide external inputs. In the above example, we have $Ra(t)$ representing the external glucose input. This external input is what we call a *perturbation*. In some cases, perturbations can help us understand how models change from their normal steady state to an alternative steady state. To investigate these properties, we need simulations of model behavior, trying out various perturbation sizes and durations.

3.4 | Measurements

To build a model, we often don't rely solely on knowledge of a specific system. Additionally, measurements from experiments can be used to create or validate a model. In this section, we will introduce various types of experiments and measurements that can be used in conjunction with building or using a computer model.

Fundamental reaction information can be measured by recreating the situation *in vitro* and measuring the reaction rates for different values of substrate, as well as possibly varying the environment. However, some processes cannot easily be translated from *in vitro* experiments to *in vivo* processes. Furthermore, some conditions may be very difficult to simulate in a test tube, such as obesity or liver disease. To perform measurements of biological systems, model organisms can be used to recreate *in vivo* conditions. For modelling metabolic systems, mouse models are frequently used, but measurements of humans in clinical trials are also common.

The specific type of measurements that can be used also influence the way models are structured. In many *in vivo* conditions, the state variables are often difficult to measure directly, and require additional processes to get an (indirect) measurement, such as fluorescence. These additional processes then also need to be taken into account in the model that is built. In other cases, we may be limited by the amount of detailed

measurements we can do, which directly limits the amount of detail we can put in our model.

A special type of measurements is done using *tracers*. These are molecules that have a radioactive or stable isotope attached to them, that can be measured afterwards. An example of the use of stable isotope tracers is the large Dalla-Man meal simulation model [9], which enabled the measurement of specific subprocesses of glucose metabolism.

3.4.1 | Parameter Estimation

When we have measurements, we also need to couple them to model parameters. If we have *in vitro* kinetic measurements, we can directly derive the kinetic parameters, for example using a Lineweaver-Burk plot for Michaelis-Menten kinetics. However, a common way to obtain the model parameters from measurements of the state variables is through parameter estimation. This procedure is beyond the scope of these lecture notes, but the general idea is that you use mathematical optimization techniques, as also used in machine learning, to select parameter values that minimize the difference between the observed state variables and the simulated state variables from the model.

3.5 | Pharmacokinetic Modeling

In this book chapter, we will explore a somewhat different application of modeling with differential equations. Instead of focusing on biological processes that occur to natural stimuli, we are now turning to models of how our bodies deal with drugs. This can be divided into two areas of research: pharmacokinetics and pharmacodynamics. The first deals with how the concentrations of drugs in our body change over time after various types of administration and dosing, while the latter involves the study of the biochemical effects of drugs. In short, pharmacokinetics studies what the body does to drugs, while pharmacodynamics focuses on what drugs do to the body. We will be discussing pharmacokinetics, as the components involved in modeling these systems resembles how we approached biological system modeling. Pharmacodynamics however, also requires an understanding of the specific chemical reactions that occur between a drug within the human body, which is beyond the scope of these lecture notes.

The main principle in pharmacokinetics revolves around the determination of the absorption, distribution, metabolism, and excretion (ADME) of drugs following administration. Applications of this range from determining the dose-response curve of different drugs, and in different physiological and pathological conditions, to determining optimal personalized drug doses.

But instead of directly turning to the components of pharmacokinetics, we will first introduce a few basic concepts that are essential to grasp before diving into the details. When administering or prescribing drugs, the main goal is that they are effective, which means that the amount of drug inside someone's system should reach a level where it can be effective. However, it cannot exceed the concentration required to achieve toxicity. Additionally, we may want to minimize the dosage initially, so we will be able to increase it without coming too close to a toxic dose. To be able to produce solutions to these problems, models can be made of the behavior of drugs inside the body. In this section, we will explore modelling concepts that are necessary for building and understanding these models.

3.5.1 | Compartmental Models

To describe the distribution of drugs after administration, compartments are used in pharmacokinetics. A compartmental model is also often used in epidemiology, for example to describe disease spread over a population. [10] Within a compartment, we assume that we have an instant homogeneous distribution of substrate. The quantity of substrate (q_i) within a compartment i , can be described according to:

$$\frac{dq_i}{dt} = \text{input}(\mathbf{q}, t) - \text{output}(\mathbf{q}, t) \quad (75)$$

Where \mathbf{q} is the vector of all masses in all compartments in the system. As opposed to earlier models, observe that in this case, the differential equation describes substrate *quantity* instead of substrate *concentration*. To convert the differential equation to substrate concentration, we will need to divide the quantity by the *volume of distribution* (V_d) of the substrate in the compartment. This is not an actual volume but it is the amount of blood that would be required if the drug was evenly distributed over the body at the concentration of the collected sample. As we assume this volume is kept constant, we can freely divide q_i by V_d within Equation 75.

One-Compartment Model

The one-compartment model is the simplest compartmental model in pharmacokinetics. As the name implies, it contains a single volume which contains the species or drug of interest. The one-compartment model is effective in describing drugs that are administered intravenously and remain in specific organs that have a high blood perfusion. This compartment typically combines the heart, liver, kidneys and the blood plasma into one *central compartment*, as seen in Figure 22.

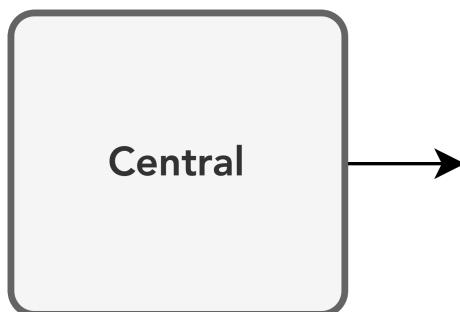


Figure 22 : One-compartment model with a central compartment containing the drug, and a single elimination term.

For an IV bolus administration of a dose D at $t = 0$, the one-compartment model equation is given as:

$$\frac{dq(t)}{dt} = -k_q q(t) \quad (76)$$

Where $q(0) = D$.

We can solve this differential equation easily by writing

$$\int \frac{dq}{q} = -k \cdot \int dt \quad (77)$$

Which results in

$$\begin{aligned}\ln(q(t)) &= \ln(D) - kt \\ \Rightarrow q(t) &= De^{-k \cdot t}\end{aligned}\tag{78}$$

An important characteristic for a drug is the elimination half-life, which is defined as the time it takes for the drug amount to become half of its initial concentration. As V_d is constant, half the concentration means half the amount of drug delivered, so we can derive a formula for this using

$$\begin{aligned}q(t_{1/2}) &= De^{-k \cdot t_{1/2}} = \frac{D}{2} \\ \Rightarrow k \cdot t_{1/2} &= \ln(2) \\ \Rightarrow t_{1/2} &= \frac{\ln(2)}{k}\end{aligned}\tag{79}$$

To test whether a drug concentration after IV bolus injection can be modelled using a one-compartment model, we can plot the log-concentration value over time and inspect whether it has a linear slope. If the points at the low and high concentration values deviate from a linear slope, this may be reason to suspect that more compartments are necessary. However, this can also occur when the measuring equipment has a lower accuracy in specific low or high concentrations, or when the measurement device nears its limit of detection, which is the lowest concentration of drug that the test can measure.

Two-Compartment Model

The most commonly used pharmacokinetic compartmental model is the two-compartment model. As the name indicates, this model contains two volumes where the drug can reside in. As in the one-compartment model, this model contains the central compartment, but it also contains a *peripheral compartment*. While the compartments in this model have no direct physiological meaning, a reasonable assumption is to think of the central compartment as the highly-perfused tissues, where the drug administered spreads rapidly, while the peripheral compartment represents the tissues with a lower perfusion rate, such as the bone or the adipose tissue.

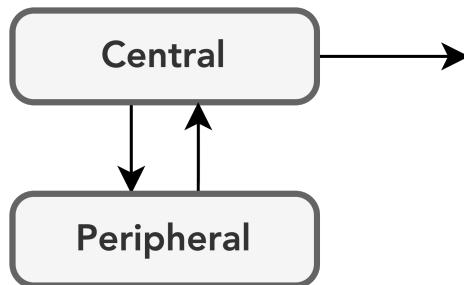


Figure 23 : Two-compartment model with a central and a peripheral compartment containing the drug, which have exchange terms and the central compartment has an elimination term.

For an IV-bolus, the ODE system of the two-compartment model is

$$\begin{aligned}\frac{dq_1(t)}{dt} &= -(k_0 + k_1)q_1(t) + k_2q_2(t) \\ \frac{dq_2(t)}{dt} &= k_1q_1(t) - k_2q_2(t)\end{aligned}\tag{80}$$

The solution of this model is more difficult, and requires the Laplace transform, which is beyond the scope of these lecture notes. Nevertheless, the solution for the central compartment is given by

$$q_1(t) = C_1 e^{-\alpha t} + C_2 e^{-\beta t} \quad (81)$$

The constants C_1 , C_2 , α and β are not in the original model equations, but can be calculated from them. α and β are known as the [macro]-rate constants. The four constants are related to the original model parameters as

$$\begin{aligned} \alpha + \beta &= k_0 + k_1 + k_2 \\ \alpha \cdot \beta &= k_2 \cdot k_0 \\ C_1 &= \frac{D_0(\alpha - k_2)}{\alpha - \beta} \\ C_2 &= \frac{D_0(k_2 - \beta)}{\alpha - \beta} \end{aligned} \quad (82)$$

Where D_0 is the initial bolus IV dose.

From the equations, one can see that the drug clearance essentially has two phases. We have the fast-phase, which is controlled by C_1 and α , and the slow phase, controlled by C_2 and β . Each of these phases also has their own half-life, which are named the distribution and elimination half-life respectively:

$$\begin{aligned} t_{1/2,\alpha} &= \frac{\ln(2)}{\alpha} \\ t_{1/2,\beta} &= \frac{\ln(2)}{\beta} \end{aligned} \quad (83)$$

The reported half life of a drug adhering to two-compartmental kinetics is often only one value, which corresponds to the slowest of the two.

Physiologically-Based Pharmacokinetic Models

In traditional pharmacokinetic models, mainly one and two-compartment models are used, with some models containing more compartments, for example when a specific tissue of interest is modelled separately. However, when more detailed information is desired, we can turn to so-called *Physiologically-based pharmacokinetic* (PBPK) models. These models contain separate compartments for many tissues in the body, and separate arterial and venous blood as compartments. Figure 24 shows a four-compartment PBPK model. Each of these compartments has its own rate equations, describing the blood flow through a tissue and the metabolism happening.

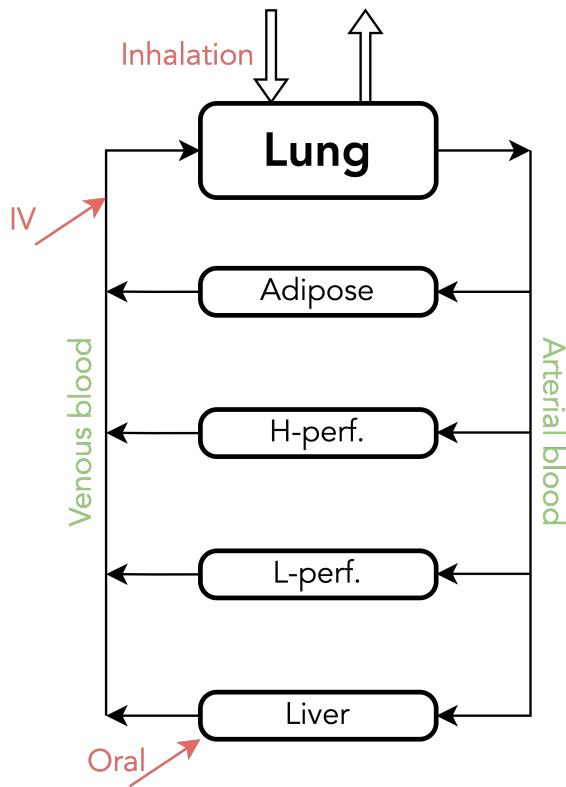


Figure 24 : A four-compartment PBPK model, showing various routes of administration, such as through inhalation, IV, or oral.

The general form of a rate equation for a molecule A in compartment x PBPK model is

$$\frac{dA_x}{dt} = Q_x \cdot \frac{A_{\text{art}} - A_x/P_x - M_x(A_x)}{V_x} \quad (84)$$

Where A_{art} is the concentration in the arterial blood, Q_x is the blood flow through compartment x , V_x is the volume of the compartment x and P_x is called the partition coefficient for compartment x , which describes how long a molecule remains in a specific compartment. A low partition coefficient means that a molecule disappears quickly from a tissue, while a high partition coefficient indicates that a molecule remains in a tissue for a longer time. Finally, $M_x(A)$ is the metabolism rate of molecule A in the specific compartment. This metabolism can for example be described using michaelis-menten kinetics, while taking into account the volume and partition coefficient of the specific organ:

$$M_x(A_x) = V_{\max} \frac{A_x/P_x}{K_m + A_x/P_x} \quad (85)$$

3.5.2 | Drug Administration

Besides compartments that describe the concentrations over time within the body, different routes of administration may also lead to variations in appearance profiles of the drugs. For example, if a drug is administered intravenously, the full dose ends up in the blood stream, while an orally dosed drug may not be fully absorbed by the gastrointestinal tract, leading to a lower absorbed dose. To ensure correct dosage and to prevent toxicity, an accurate representation of drug appearance for different administration routes is critical. In this section, we'll explore the modelling of these routes of administration.

3.5.3 | Metabolism and Excretion

The First-Pass Effect

The Blood-Brain Barrier

3.5.4 | Outcome Measures

3.5.5 | Examples of Pharmacokinetic Models

$$\left(\begin{array}{cccc|c} A & B & C & D & \\ \hline 0 & 2 & 0 & 0 & A \\ 2 & 8 & 0 & 1 & B \\ 0 & 0 & 0 & 4 & C \\ 0 & 1 & 4 & 0 & D \end{array} \right) \quad (86)$$

$$\left(\begin{array}{cccc|c} A & B & C & D & \\ \hline 0 & 1 & 0 & 0 & A \\ 1 & 2 & 0 & 1 & B \\ 0 & 0 & 0 & 1 & C \\ 0 & 1 & 1 & 0 & D \end{array} \right) \quad (87)$$

3.6 | Exercises

Bibliography

- [1] A. Aderem, "Systems biology: Its practice and challenges," *Cell*, vol. 121, no. 4, pp. 511–513, 2005, doi: [10.1016/j.cell.2005.04.020](https://doi.org/10.1016/j.cell.2005.04.020).
- [2] E. O. Voit, "Perspective: Systems biology beyond biology," *Frontiers in Systems Biology*, vol. 2, p. 987135–987136, 2022, doi: [10.3389/FSYSB.2022.987135](https://doi.org/10.3389/FSYSB.2022.987135).
- [3] N. A. van Riel, "Parameter Estimation and Uncertainty Analysis for Biological Systems," 2020.
- [4] R. Diestel, "The Basics," 5th ed., Springer-Verlag GmbH, 2017, pp. 1–34. doi: [10.1007/978-3-662-53622-3_1](https://doi.org/10.1007/978-3-662-53622-3_1).
- [5] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A gene-coexpression network for global discovery of conserved genetic modules," *Science*, vol. 302, no. 5643, pp. 249–255, 2003, doi: [10.1126/SCIENCE.1087447/SUPPL_FILE/STUART.PDF](https://doi.org/10.1126/SCIENCE.1087447).
- [6] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature Communications* 2019 10:1, vol. 10, no. 1, pp. 1–10, 2019, doi: [10.1038/s41467-019-08746-5](https://doi.org/10.1038/s41467-019-08746-5).
- [7] A. Goldbeter, "A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 20, 1991, doi: [10.1073/pnas.88.20.9107](https://doi.org/10.1073/pnas.88.20.9107).
- [8] R. N. Bergman, "Origins and History of the Minimal Model of Glucose Regulation," *Frontiers in endocrinology*, vol. 11, 2021, doi: [10.3389/FENDO.2020.583016](https://doi.org/10.3389/FENDO.2020.583016).
- [9] C. D. Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1740–1749, 2007, doi: [10.1109/TBME.2007.893506](https://doi.org/10.1109/TBME.2007.893506).
- [10] F. Brauer, "Compartmental Models in Epidemiology," *Lecture Notes in Mathematics*, vol. 1945, pp. 19–79, 2008, doi: [10.1007/978-3-540-78911-6_2](https://doi.org/10.1007/978-3-540-78911-6_2).